

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number  
**WO 03/066824 A2**

- (51) International Patent Classification<sup>7</sup>: **C12N**
- (21) International Application Number: PCT/US03/03616
- (22) International Filing Date: 7 February 2003 (07.02.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/355,547 7 February 2002 (07.02.2002) US
- (71) Applicants (*for all designated States except US*): **AVENTIS BEHRING GMBH** [DE/DE]; Emil-von-Behring-Strasse 76, 35041 Marburg (DE). **AVENTIS BEHRING L.L.C.** [US/US]; 1020 First Avenue, King of Prussia, PA 19406 (US). **DELTA BIOTECHNOLOGY LIMITED** [GB/GB]; Castle Court, 59 Castle Boulevard, Nottingham NG7 1FD (GB). **DYAX CORPORATION** [US/US]; 300 Technology Square, Cambridge, MA 02139 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **HAUSER, Hans-Peter** [DE/DE]; Feuerdornweg 6, 35041 Marburg (DE).

**WEIMER, Thomas** [DE/DE]; Richard-Wagner-Strasse 8, 35075 Gladen Bach (DE). **ROMBERG, Val** [US/US]; 1330 Old Schuykill Road, Parkerford, PA 19457 (US). **KEE, Scott, M.** [US/US]; 218 Kathy Drive, Bourbonnais, IL 60914 (US). **SLEEP, Darrell** [GB/GB]; 66 Ladybag Road, West Bridgford Boulevard, Nottingham NG2 5DS (GB). **LADNER, Robert, Charles** [US/US]; 3827 Green Valley Road, Ijamsville, MD 21754 (US). **LEY, Arthur, C.** [US/US]; 122 Adena Road, Newton, MA 02465 (US).

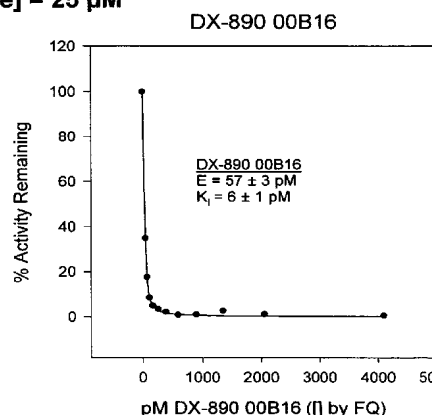
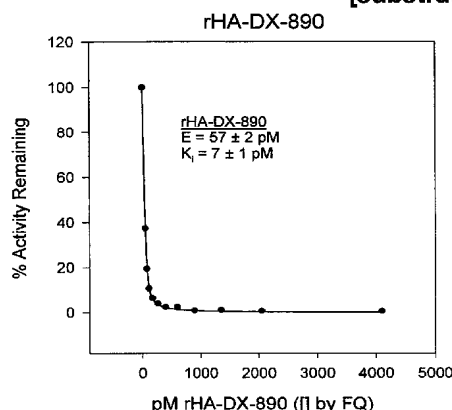
(74) Agent: **MORRY, Mary, J.**; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

[Continued on next page]

(54) Title: ALBUMIN-FUSED KUNITZ DOMAIN PEPTIDES

rHA-DX-890 Batch 1743#09 K<sub>i</sub> Determination[HNE] = 100 pM  
[substrate] = 25 μM

(57) Abstract: The invention relates to proteins comprising serine protease inhibiting peptides, such as Kunitz domain peptides (including, but not limited to, fragments and variants thereof) fused to albumin, or fragments or variants thereof. These fusion proteins are herein collectively referred to as "albumin fusion proteins of the invention." These fusion proteins exhibit extended shelf-life and/or extended or therapeutic activity in solution. The invention encompasses, therapeutic albumin fusion proteins, compositions, pharmaceutical compositions, formulations and kits. The invention also encompasses nucleic acid molecules encoding the albumin fusion proteins of the invention, as well as vectors containing these nucleic acids, host cells transformed with these nucleic acids and vectors, and methods of making the albumin fusion proteins of the invention using these nucleic acids, vectors, and/or host cells. The invention also relates to compositions and methods for inhibiting neutrophil elastase, kallikrein, and plasmin. The invention further relates to compositions and methods for treating cystic fibrosis and cancer.



WO 03/066824 A2



Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI,  
SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN,  
GQ, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

— *without international search report and to be republished  
upon receipt of that report*

## **Albumin-Fused Kunitz Domain Peptides**

### **Related Applications**

This application claims priority to U.S. Provisional Application Serial No. 60/355,547, filed February 7, 2002. The disclosure of that application is incorporated herein by reference in its entirety.

### **Field of the Invention**

The invention relates to the fields of Kunitz domain peptides and albumin fusion proteins. More specifically, the invention relates to Kunitz domain peptides and albumin fusion proteins for treating, preventing, or ameliorating a disease or disorder.

### **Background of the Invention**

A Kunitz domain is a folding domain of approximately 51-64 residues which forms a central anti-parallel beta sheet and a short C-terminal helix (see e.g., U.S. Patent No. 6,087,473, which is hereby incorporated by reference in its entirety). This characteristic domain comprises six cysteine residues that form three disulfide bonds, resulting in a double-loop structure. Between the N-terminal region and the first beta strand resides the active inhibitory binding loop. This binding loop is disulfide bonded through the P2 C<sub>14</sub> residue to the hairpin loop formed between the last two beta strands. Isolated Kunitz domains from a variety of proteinase inhibitors have been shown to have inhibitory activity (e.g., Petersen et al., Eur. J. Biochem. 125:310-316, 1996; Wagner et al., Biochem. Biophys. Res. Comm. 186:1138-1145, 1992; Dennis et al., J. Biol. Chem. 270:25411-25417, 1995).

Linked Kunitz domains also have been shown to have inhibitory activity, as discussed, for example, in U.S. Patent No. 6,087,473. Proteinase inhibitors comprising one or more Kunitz domains include tissue factor pathway inhibitor (TFPI), tissue factor pathway inhibitor 2 (TFPI-2), amyloid  $\beta$ -protein precursor (A $\beta$ PP), aprotinin, and placental bikunin. TFPI, an extrinsic pathway inhibitor and a natural anticoagulant, contains three tandemly linked Kunitz inhibitor domains. The amino-terminal Kunitz domain inhibits factor VIIa, plasmin, and cathepsin G; the second domain inhibits factor Xa, trypsin, and chymotrypsin; and the third domain has no known activity (Petersen et al., *ibid.*).

The inhibitory activity of Kunitz domain peptides towards serine proteases has been demonstrated in several previous studies. The following subsections discuss studies of the

- 2 -

inhibition of serine proteases, such as plasma kallikrein, plasmin, and neutrophil elastase by Kunitz Domain peptides.

### **Plasma Kallikrein Inhibitors**

Kallikreins are serine proteases found in both tissues and plasma [see, for example, U.S. Patent No. 6,333,402 to Markland, which is hereby incorporated by reference in its entirety]. Plasma kallikrein is involved in contact-activated (intrinsic pathway) coagulation, fibrinolysis, hypotension, and inflammation [See Bhoola, K.D., C.D. Figueroa, and K. Worthy, Pharmacological Reviews (1992) 44(1)1-80]. These effects of kallikrein are mediated through the activities of three distinct physiological substrates:

- i) Factor XII (coagulation),
- ii) Pro-urokinase/plasminogen (fibrinolysis), and
- iii) Kininogens (hypotension and inflammation).

Kallikrein cleavage of kininogens results in the production of kinins, small highly potent bioactive peptides. The kinins act through cell surface receptors, designated BK-1 and BK-2, present on a variety of cell types including endothelia, epithelia, smooth muscle, neural, glandular and hematopoietic. Intracellular heterotrimeric G-proteins link the kinin receptors to second messenger pathways including nitric oxide, adenylyl cyclase, phospholipase A<sub>2</sub> and phospholipase C. Among the significant physiological activities of kinins are: (i) increased vascular permeability; (ii) vasodilation; (iii) bronchospasm; and (iv) pain induction. Thus, kinins mediate the life-threatening vascular shock and edema associated with bacteremia (sepsis) or trauma, the edema and airway hyperreactivity of asthma, and both inflammatory and neurogenic pain associated with tissue injury. The consequences of inappropriate plasma kallikrein activity and resultant kinin production are dramatically illustrated in patients with hereditary angioedema (HAE). HAE is due to a genetic deficiency of C1-inhibitor, the principal endogenous inhibitor of plasma kallikrein. Symptoms of HAE include edema of the skin, subcutaneous tissues and gastrointestinal tract, and abdominal pain and vomiting. Nearly one-third of HAE patients die by suffocation due to edema of the larynx and upper respiratory tract. Kallikrein is secreted as a zymogen (prekallikrein) that circulates as an inactive molecule until activated by a proteolytic event. [Genebank entry P03952 shows Human Plasma Prekallikrein.]

An important inhibitor of plasma kallikrein (pKA) in vivo is the C1 inhibitor; (see Schmaier, et al. in "Contact Activation and Its Abnormalities", Chapter 2 in Hemostasis and Thrombosis, Colman, R W, J Hirsh, V J Marder, and E W Salzman, Editors, Second Edition,

- 3 -

1987, J. B. Lippincott Company, Philadelphia, PA., pp.27-28). C1 is a serpin and forms an irreversible or nearly irreversible complex with pKA. Although bovine pancreatic trypsin inhibitor (also known as BPTI, aprotinin, or Trasylol<sup>TM</sup>) was initially thought to be a strong pKA inhibitor with  $K_i = 320$  pM [Auerswald, E.-A., D. Hoerlein, G. Reinhardt, W. Schroder, and E. Schnabel, Bio. Chem. Hoppe-Seyler, (1988), 369 (Supplement):27-35], a more recent report [Berndt, et al., Biochemistry, 32:4564-70, 1993] indicates that its  $K_i$  for plasma Kallikrein is 30 nM (i.e., 30,000 pM). The G36S mutant had a  $K_i$  of over 500 nM.

Markland et al. [U.S. Patent Nos. 6,333,402; 5,994,125; 6,057,287; and 5,795,865; each reference hereby incorporated by reference in its entirety] claim a number of derivatives having high affinity and specificity in inhibiting human plasma kallikrein. One of these proteins is being tested in human patients who have HAE. Although early indications are that the compound is safe and effective, the duration of effect is shorter than desired.

### **Plasmin Inhibitors**

Plasmin is a serine protease derived from plasminogen. The catalytic domain of plasmin (or "CatDom") cuts peptide bonds, particularly after arginine residues and to a lesser extent after lysines and is highly homologous to trypsin, chymotrypsin, kallikrein, and many other serine proteases. Most of the specificity of plasmin derives from the kringles' binding of fibrin (Lucas et al., J Biological Chem (1983) 258(7)4249-56.; Varadi & Patthy, Biochemistry (1983) 22:2440-2446.; and Varadi & Patthy, Biochemistry (1984) 23:2108-2112.). On activation, the bond between ARG<sub>561</sub> -Val<sub>562</sub> is cut, allowing the newly free amino terminus to form a salt bridge. The kringles remain, nevertheless, attached to the CatDom through two disulfides (Colman, R W, J Hirsh, V J Marder, and E W Salzman, Editors, Hemostasis and Thrombosis, Second Edition, 1987, J. B. Lippincott Company, Philadelphia, Pa., Bobbins, 1987, *supra*).

The agent mainly responsible for fibrinolysis is plasmin the activated form of plasminogen. Many substances can activate plasminogen, including activated Hageman factor, streptokinase, urokinase (uPA), tissue-type plasminogen activator (tPA), and plasma kallikrein (pKA). pKA is both an activator of the zymogen form of urokinase and a direct plasminogen activator.

Plasmin is undetectable in normal circulating blood, but plasminogen, the zymogen, is present at about 3  $\mu$ M. An additional, unmeasured amount of plasminogen is bound to fibrin and other components of the extracellular matrix and cell surfaces. Normal blood contains the physiological inhibitor of plasmin,  $\alpha_2$ -plasmin inhibitor ( $\alpha_2$ -PI), at about 2  $\mu$ M. Plasmin and

- 4 -

$\alpha_2$  -PI form a 1:1 complex. Matrix or cell bound-plasmin is relatively inaccessible to inhibition by  $\alpha_2$  -PI. Thus, activation of plasmin can exceed the neutralizing capacity of  $\alpha_2$  -PI causing a profibrinolytic state.

Plasmin, once formed:

- i) degrades fibrin clots, sometimes prematurely;
- ii) digests fibrinogen (the building material of clots) impairing hemostasis by causing formation of friable, easily lysed clots from the degradation products, and inhibition of platelet adhesion/aggregation by the fibrinogen degradation products;
- iii) interacts directly with platelets to cleave glycoproteins Ib and IIb/IIIa preventing adhesion to injured endothelium in areas of high shear blood flow and impairing the aggregation response needed for platelet plug formation (Adelman et al., Blood (1986) 68(6)1280-1284.);
- iv) proteolytically inactivates enzymes in the extrinsic coagulation pathway further promoting a prolytic state. Robbins (Robbins, Chapter 21 of Hemostasis and Thrombosis, Colman, R. W., J. Hirsh, V. J. Marder, and E. W. Salzman, Editors, Second Edition, 1987, J. B. Lippincott Company, Philadelphia, PA) reviewed the plasminogen-plasmin system in detail. This publication (i.e., Colman, R. W., J. Hirsh, V. J. Marder, and E. W. Salzman, Editors, Hemostasis and Thrombosis, Second Edition, 1987, J. B. Lippincott Company, Philadelphia, PA) is hereby incorporated by reference.

#### *Fibrinolysis and Fibrinogenolysis*

Inappropriate fibrinolysis and fibrinogenolysis leading to excessive bleeding is a frequent complication of surgical procedures that require extracorporeal circulation, such as cardiopulmonary bypass, and is also encountered in thrombolytic therapy and organ transplantation, particularly liver. Other clinical conditions characterized by high incidence of bleeding diathesis include liver cirrhosis, amyloidosis, acute promyelocytic leukemia, and solid tumors. Restoration of hemostasis requires infusion of plasma and/or plasma products, which risks immunological reaction and exposure to pathogens, e.g. hepatitis virus and HIV.

Very high blood loss can resist resolution even with massive infusion. When judged life-threatening, the hemorrhage is treated with antifibrinolytics such as  $\epsilon$ -amino caproic acid (See Hoover et al., Biochemistry (1993) 32:10936-43) (EACA), tranexarnic acid, or aprotinin

- 5 -

(Neuhaus et al., Lancet (1989) 2(8668)924-5). EACA and tranexamic acid only prevent plasmin from binding fibrin by binding the kringles, thus leaving plasmin as a free protease in plasma. BPTI is a direct inhibitor of plasmin and is the most effective of these agents. Due to the potential for thrombotic complications, renal toxicity and, in the case of BPTI, immunogenicity, these agents are used with caution and usually reserved as a "last resort" (Putterman, Acta Chir Scand (1989) 155(6-7)367). All three of the antifibrinolytic agents lack target specificity and affinity and interact with tissues and organs through uncharacterized metabolic pathways. The large doses required due to low affinity, side effects due to lack of specificity and potential for immune reaction and organ/tissue toxicity augment against use of these antifibrinolytics prophylactically to prevent bleeding or as a routine postoperative therapy to avoid or reduce transfusion therapy. Thus, there is a need for a safe antifibrinolytic. The essential attributes of such an agent are:

- i) Neutralization of relevant target fibrinolytic enzyme(s);
- ii) High affinity binding to target enzymes to minimize dose;
- iii) High specificity for target, to reduce side effects; and
- iv) High degree of similarity to human protein to minimize potential immunogenicity and organ/tissue toxicity.

All of the fibrinolytic enzymes that are candidate targets for inhibition by an efficacious antifibrinolytic are chymotrypsin-homologous serine proteases.

### *Excessive Bleeding*

Excessive bleeding can result from deficient coagulation activity, elevated fibrinolytic activity, or a combination of the two conditions. In most bleeding diatheses one must control the activity of plasmin. The clinically beneficial effect of BPTI in reducing blood loss is thought to result from its inhibition of plasmin ( $K_i \sim 0.3$  nM) or of plasma kallikrein ( $K_i \sim 100$  nM) or both enzymes.

Gardell [Toxicol. Pathol. (1993) 21(2)190-8] has reviewed currently-used thrombolytics, and has stated that, although thrombolytic agents (e.g. tPA) do open blood vessels, excessive bleeding is a serious safety issue. Although tPA and streptokinase have short plasma half lives, the plasmin they activate remains in the system for a long time and, as stated, the system is potentially deficient in plasmin inhibitors. Thus, excessive activation of plasminogen can lead to a dangerous inability to clot and injurious or fatal hemorrhage. A potent, highly specific plasmin inhibitor would be useful in such cases.

- 6 -

BPTI is a potent plasmin inhibitor. However, it has been found that it is sufficiently antigenic that second uses require skin testing. Furthermore, the doses of BPTI required to control bleeding are quite high and the mechanism of action is not clear. Some say that BPTI acts on plasmin while others say that it acts by inhibiting plasma kallikrein. Fraedrich et al. [Thorac Cardiovasc Surg (1989) 37(2)89-91] report that doses of about 840 mg of BPTI to 80 open-heart surgery patients reduced blood loss by almost half and the mean amount transfused was decreased by 74%. Miles Inc. has recently introduced Trasylol™ in the U.S. for reduction of bleeding in surgery [see Miles product brochure on Trasylol™, which is hereby incorporated by reference]. Lohmann and Marshal [Refract Corneal Surg (1993) 9(4)300-2] suggest that plasmin inhibitors may be useful in controlling bleeding in surgery of the eye. Sheridan et al. [Dis Colon Rectum (1989) 32(6)505-8] reports that BPTI may be useful in limiting bleeding in colonic surgery.

A plasmin inhibitor that is approximately as potent as BPTI or more potent but that is almost identical to a human protein domain offers similar therapeutic potential but poses less potential for antigenicity.

#### *Angiogenesis:*

Plasmin is the key enzyme in angiogenesis. O'Reilly et al. [Cell (1994) 79:315-328] reports that a 38 kDa fragment of plasmin (lacking the catalytic domain) is a potent inhibitor of metastasis, indicating that inhibition of plasmin could be useful in blocking metastasis of tumors [Fidler & Ellis, Cell (1994) 79:185-188; See also Ellis et al., Ann NY Acad Sci (1992) 667:13-31; O'Reilly et al., Fidler & Ellis, and Ellis et al. are hereby incorporated by reference].

#### **Neutrophil Elastase Inhibition**

Cystic Fibrosis is a hereditary, autosomal recessive disorder affecting pulmonary, gastrointestinal, and reproductive systems. With a prevalence of 80,000 worldwide, the incidence of CF is estimated at 1 in 3500 [Cystic Fibrosis Foundation, *Patient Registry 1998 Annual Data Report*, Bethesda, Maryland, September 1999]. The genetic defect in CF was described in 1989 as the loss of a single phenylalanine at position 508 ( $\Delta F508$ ), resulting in a faulty cystic fibrosis transmembrane conductance regulator protein (CFTR) which inhibits the reabsorption of  $Cl^-$  (and hence  $Na^+$  and water) [Rommens, J.M., *et al.*, "Identification of the cystic fibrosis gene: chromosome walking and jumping," *Science* 245:1059, 1989; Riordan, J.R., *et al.*, "Identification of the cystic fibrosis gene: cloning and complementary DNA,"



- 7 -

*Science* 245:1066, 1989; Kerem, B., *et al.*, "Identification of the cystic fibrosis gene: genetic analysis, *Science* 245:1073, 1989]. Mutations other than  $\Delta F508$  have been found in CFTR and may cause CF. Desiccated mucus then plugs many of the passageways in the respiratory, gastrointestinal, and reproductive systems.

More than 75% of the mortality from CF is due to respiratory complications [Cystic Fibrosis Foundation, *Patient Registry 1998 Annual Data Report*, Bethesda, Maryland, September 1999]. Although disease of the pancreas, liver, and intestine is present in CF individuals before birth, the CF lung is normal at birth and until the onset of infection and inflammation. Then, defective  $Cl^-$  reabsorption in the CF lung leads to desiccated airway secretions by drawing sodium out of the airways, with water following passively. Desiccated secretions may then interfere with mucociliary clearance by trapping bacteria in an environment well suited to colonization with distinctive microbial pathogens [Reynolds, H.Y., *et al.*, "Muroid *Pseudomonas aeruginosa*: a sign of cystic fibrosis in young adults with chronic pulmonary disease," *J.A.M.A.* 236:2190, 1976]. The ensuing lung infection and inflammation recruits and activates neutrophils which release neutrophil elastase (NE). The neutrophil-dominated inflammation on the respiratory epithelial surface results in a chronic epithelial burden of neutrophil elastase. Endogenous antiprotease is rapidly overwhelmed by an excess of NE in the CF lung. In addition, NE stimulates the production of pro-inflammatory mediators and cleaves complement receptors and IgG, thereby crippling host defense mechanisms preventing further bacterial colonization [Tosi, M.F., *et al.*, "Neutrophil elastase cleaves C3bi on opsonized *Pseudomonas* as well as CR1 on neutrophils to create a functionally important opsonin receptor mismatch," *J. Clin. Invest.* 86:300, 1990]. The infection thereby becomes persistent, and the massive ongoing inflammation and excessive levels of NE destroy the airway epithelium, leading to bronchiectasis, and the progressive loss of pulmonary function and death.

One therapeutic approach in patients with CF is the eradication of CF pathogens by systemic antimicrobials such as tobramycin and ciprofloxacin. While these specific antimicrobial agents have been shown to be effective in clearing infection and improving pulmonary function, antibiotic resistance to tobramycin and ciprofloxacin is reported in 7.5% and 9.6% of CF patients respectively [Cystic Fibrosis Foundation, *Patient Registry 1998 Annual Data Report*, Bethesda, Maryland, September 1999]. As the use of these antimicrobials for CF increases in patients of whom 60% are infected with *P. aeruginosa* and 41% with *S. aureus*, drug resistance selection pressure has increased.

- 8 -

Pulmonary function also has been a therapeutic target in patients with CF. Pulmozyme® (dornase alfa), a recombinant human deoxyribonuclease which reduces mucus viscoelasticity by hydrolyzing DNA in sputum, has been shown in clinical studies to increase FEV<sub>1</sub> and FVC after 8 days of treatment. This change last for six months, and is accompanied by a reduction in the use of intravenous antibiotics [Fuchs, H.L., *et al.*, "Effect of aerosolized recombinant human Dnase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis," *N. Engl. J. Med.*, 331:637-642, 1994].

Another therapeutic approach is to use a protease inhibitor to ablate the direct effect of NE on elastase degradation and its sequelae. Neutralization of excess NE can restore normal homeostatic balance which protects the extracellular lung matrix. Normalized antiprotease activity in the lung preserves elastin, reduces mucus viscosity through reduction of the neutrophil response, and preserves of pulmonary function, thus reducing mortality in CF. In addition, the restoration of complement-mediated phagocytosis can enable the immune system to clear bacterial pathogens, resulting in reduction of the incidence, duration, and severity of pulmonary infection. For example, in a rat model of CF, after seven days of treatment with alpha<sub>1</sub> antitrypsin reduced bacterial counts to  $0.2 \pm 0.4$ , compared to  $85 \pm 21$  in the placebo group [Cantin, A. and Woods, D, "Aerosolized Prolastin Suppresses Bacterial Proliferation in a Model of Chronic *Pseudomonas aeruginosa* Lung Infection" *Am J Respir Crit Care Med* 160:1130-1136, 1999]

### **Summary of the Invention**

The invention relates to proteins comprising Kunitz domain peptides fused to albumin. These fusion proteins are herein collectively referred to as "albumin fusion proteins of the invention." These fusion proteins of the invention exhibit extended *in vivo* half-life and/or extended or therapeutic activity in solution.

The invention encompasses therapeutic albumin fusion proteins, compositions, pharmaceutical compositions, formulations and kits. The invention also encompasses nucleic acid molecules encoding the albumin fusion proteins of the invention, as well as vectors containing these nucleic acids, host cells transformed with these nucleic acids and vectors, and methods of making the albumin fusion proteins of the invention using these nucleic acids, vectors, and/or host cells.

An object of the invention is to provide an albumin fusion protein comprising a Kunitz domain peptide or a fragment or variant thereof, and albumin, or a fragment or variant thereof. Suitable Kunitz domain peptides for use in such albumin fusion proteins include DX-890,

- 9 -

DX-88, DX-1000, and DPI-14. The Kunitz domain peptide portion optionally may be separated from the albumin portion by a linker. Another object of the invention is to provide compositions and methods involving albumin fusion proteins for inhibiting serine proteases, non-limiting examples of which include plasma kallikrein, plasmin and neutrophil elastase.

Another aspect of the invention is to provide an albumin fusion protein comprising at least two Kunitz domain peptides or fragments or variants thereof, wherein at least one of the Kunitz domain peptide or fragment or variant has a functional activity, such as inhibiting plasmin, kallikrein, or human neutrophil elastase.

Yet another aspect of this invention is to provide an albumin fusion protein comprising a Kunitz domain peptide, or a fragment or variant thereof, and albumin, or a fragment or variant thereof, wherein the albumin has an albumin activity that prolongs the *in vivo* half-life of a Kunitz domain peptide, such as DX-890, DX-88, DX-1000, and DPI-14, or a fragment or variant thereof, compared to the *in vivo* half-life of the Kunitz domain peptide or a fragment or variant thereof in an unfused state.

Yet another aspect of this invention is to provide an albumin fusion protein comprising a Kunitz domain peptide, or a fragment or variant thereof, and albumin, or a fragment or variant thereof, wherein the albumin fusion protein of the invention has increased solubility at physiological pH.

One aspect of the invention is to provide an albumin fusion protein comprising a Kunitz domain peptide, or fragment or variant thereof, and albumin, or fragment or variant thereof, wherein the Kunitz domain peptide, or fragment or variant thereof, is fused to the N-terminus of albumin or to the N-terminus of the fragment or variant of albumin. Alternatively, this invention also provides an albumin fusion protein comprising a Kunitz domain peptide, or fragment or variant thereof, and albumin, or fragment or variant thereof, wherein the Kunitz domain peptide, or fragment or variant thereof, is fused to the C-terminus of albumin or to the C-terminus of the fragment or variant of albumin.

This invention provides a composition comprising an albumin fusion protein and a pharmaceutically acceptable carrier. Another object of the invention is to provide a method of treating a patient with cystic fibrosis, a cystic fibrosis-related disease or disorder, or a disease or disorder that can be modulated by a Kunitz domain peptide comprising DX-890 and/or DPI-14. The method comprises the step of administering an effective amount of the albumin fusion protein comprising a Kunitz domain peptide that comprises DX-890 and/or DPI-14, or fragment or variant thereof, and albumin, or fragment or variant thereof.

- 10 -

Another object of this invention is to provide a method of treating a patient with hereditary angioedema, a hereditary angioedema-related disease or disorder, or a disease that is modulated by a Kunitz domain peptide such as DX-88. The method comprises the step of administering an effective amount of the albumin fusion protein, wherein the albumin fusion protein comprises a Kunitz domain peptide comprising DX-88, or fragment or variant thereof, and albumin, or fragment or variant thereof.

An object of this invention is to provide a method of treating a patient with cancer, a cancer-related disease, bleeding, or disease that is modulated by a Kunitz domain peptide such as DX-1000. The method comprises the step of administering an effective amount of the albumin fusion protein, wherein the albumin fusion protein comprises a Kunitz domain peptide comprising DX-1000, or fragment or variant thereof, and albumin, or fragment or variant thereof.

Another object of the invention is to provide a nucleic acid molecule comprising a polynucleotide sequence encoding an albumin fusion protein, as well as a vector that comprises such a nucleic acid molecule.

The invention also provides a method for manufacturing a albumin fusion protein, wherein the method comprises:

- (a) providing a nucleic acid comprising a nucleotide sequence encoding the albumin fusion protein expressible in an organism;
- (b) expressing the nucleic acid in the organism to form an albumin fusion protein; and
- (c) purifying the albumin fusion protein.

#### **Brief Description of the Drawings**

Figure 1:  $K_i$  measurements of DX-890 and the DX-890-HSA fusion.

Figure 2: Plasma clearance curves for  $^{125}\text{I}$ -DX-890 (left) and  $^{125}\text{I}$ -DX-890-HSA fusion (right).

Figure 3:  $^{125}\text{I}$ -DX890 in normal mouse plasma on SE-HPLC (Superose-12).

Figure 4: SE-HPLC(Superose-12) Profiles of  $^{125}\text{I}$ -HAS-DX890 in normal mouse plasma..

Figure 5: Plasma Clearance of  $^{125}\text{I}$  Labeled DX-890 and HSA-DX-890 in Rabbits

Figure 6: SEC Analysis of Rabbit Plasma Samples

**Detailed Description of the Invention**

The present invention relates to albumin-fused Kunitz domain peptides. The present invention also relates to bifunctional (or multifunctional) fusion proteins in which albumin is coupled to two (or more) Kunitz domain peptides, optionally different Kunitz domain peptides. Such bifunctional (or multifunctional) fusion proteins having different Kunitz domain peptides are expected to have an improved drug resistance profile as compared to an albumin fusion protein comprising only one type of Kunitz domain peptide. Some conditions may require inhibition of two or more proteases and fusion of multiple Kunitz domains allows one compound to be used for inhibition of the two or more proteases. Alternatively, one can fuse two or more Kunitz domains, each directed to the same protease so that the inhibitor activity per gram is increased. A useful form of an inhibitor having two Kunitz domains is  $K_1::SA::K_2$ , where  $K_1$  and  $K_2$  are the Kunitz domains and SA is serum albumin or a substantial portion thereof. Such bifunctional (or multifunctional) fusion proteins may also exhibit synergistic effects, as compared to an albumin fusion protein comprising only one type of Kunitz domain peptide. Furthermore, chemical entities may be covalently attached to the fusion proteins of the invention to enhance a biological activity or to modulate a biological activity.

The albumin fusion proteins of the present invention are expected to prolong the half-life of the Kunitz domain peptide *in vivo*. The *in vitro* or *in vivo* half-life of said albumin-fused peptide is extended 2-fold, or 5-fold, or more, over the half-life of the peptide lacking the linked albumin. Furthermore, due at least in part to the increased half-life of the peptide, the albumin fusion proteins of the present invention are expected to reduce the frequency of the dosing schedule of the therapeutic peptide. The dosing schedule frequency is reduced by at least one-quarter or by at least one-half, as compared to the frequency of the dosing schedule of the therapeutic peptide lacking the linked albumin.

The albumin fusion proteins of the present invention prolong the shelf life of the peptide, and/or stabilize the peptide and/or its activity in solution (or in a pharmaceutical composition) *in vitro* and/or *in vivo*. These albumin fusion proteins, which may be therapeutic agents, are expected to reduce the need to formulate protein solutions with large excesses of carrier proteins (such as albumin, unfused) to prevent loss of proteins due to factors such as nonspecific binding.

The present invention also encompasses nucleic acid molecules encoding the albumin fusion proteins as well as vectors containing these nucleic acids, host cells transformed with these nucleic acids vectors, and methods of making the albumin fusion proteins of the

- 12 -

invention using these nucleic acids, vectors, and/or host cells. The present invention further includes transgenic organisms modified to contain the nucleic acid molecules of the invention, optionally modified to express the albumin fusion proteins encoded by the nucleic acid molecules.

### **Albumin**

The terms, human serum albumin (HSA) and human albumin (HA) are used interchangeably herein. The terms, "albumin" and "serum albumin" are broader, and encompass human serum albumin (and fragments and variants thereof) as well as albumin from other species (and fragments and variants thereof).

As used herein, "albumin" refers collectively to albumin protein or amino acid sequence, or an albumin fragment or variant, having one or more functional activities (e.g., biological activities) of albumin. In particular, "albumin" refers to human albumin or fragments thereof (see EP 201 239, EP 322 094 WO 97/24445, WO95/23857) especially the mature form of human albumin as shown in SEQ ID NO:18 herein and in Table 1 and SEQ ID NO:18 of U.S. Provisional Application Serial No. 60/355,547 and WO 01/79480 or albumin from other vertebrates or fragments thereof, or analogs or variants of these molecules or fragments thereof.

The human serum albumin protein used in the albumin fusion proteins of the invention contains one or both of the following sets of point mutations with reference to SEQ ID NO:18: Leu-407 to Ala, Leu-408 to Val, Val-409 to Ala, and Arg-410 to Ala; or Arg-410 to Ala, Lys-413 to Gln, and Lys-414 to Gln (see, e.g., International Publication No. WO95/23857, hereby incorporated in its entirety by reference herein). In some embodiments, albumin fusion proteins of the invention that contain one or both of above-described sets of point mutations have improved stability/resistance to yeast Yap3p proteolytic cleavage, allowing increased production of recombinant albumin fusion proteins expressed in yeast host cells.

As used herein, a portion of albumin sufficient to prolong or extend the *in vivo* half-life, therapeutic activity, or shelf-life of the Therapeutic protein refers to a portion of albumin sufficient in length or structure to stabilize, prolong or extend the *in vivo* half-life, therapeutic activity or shelf life of the Therapeutic protein portion of the albumin fusion protein compared to the *in vivo* half-life, therapeutic activity, or shelf-life of the Therapeutic protein in the non-fusion state. The albumin portion of the albumin fusion proteins may comprise the full length of the HA sequence as described above, or may include one or more fragments thereof that are

- 13 -

capable of stabilizing or prolonging the therapeutic activity. Such fragments may be of 10 or more amino acids in length or may include about 15, 20, 25, 30, 50, or more contiguous amino acids from the HA sequence or may include part or all of specific domains of HA.

The albumin portion of the albumin fusion proteins of the invention may be a variant of normal HA. The Therapeutic protein portion of the albumin fusion proteins of the invention may also be variants of the Therapeutic proteins as described herein. The term "variants" includes insertions, deletions and substitutions, either conservative or non-conservative, where such changes do not substantially alter one or more of the oncotic, useful ligand-binding and non-immunogenic properties of albumin, or the active site, or active domain which confers the therapeutic activities of the Therapeutic proteins.

In particular, the albumin fusion proteins of the invention may include naturally occurring polymorphic variants of human albumin and fragments of human albumin, for example those fragments disclosed in EP 322 094 (namely HA (P<sub>n</sub>), where n is 369 to 419). The albumin may be derived from any vertebrate, especially any mammal, for example human, cow, sheep, or pig. Non-mammalian albumins include, but are not limited to, hen and salmon. The albumin portion of the albumin fusion protein may be from a different animal than the Therapeutic protein portion.

Generally speaking, an HA fragment or variant will be at least 100 amino acids long, for example, at least 150 amino acids long. The HA variant may consist of or alternatively comprise at least one whole domain of HA, for example domains 1 (amino acids 1-194 of SEQ ID NO:18), 2 (amino acids 195-387 of SEQ ID NO:18), 3 (amino acids 388-585 of SEQ ID NO:18), 1 + 2 (1-387 of SEQ ID NO:18), 2 + 3 (195-585 of SEQ ID NO:18) or 1 + 3 (amino acids 1-194 of SEQ ID NO:18+ amino acids 388-585 of SEQ ID NO:18). Each domain is itself made up of two homologous subdomains namely 1-105, 120-194, 195-291, 316-387, 388-491 and 512-585, with flexible inter-subdomain linker regions comprising residues Lys106 to Glu119, Glu292 to Val315 and Glu492 to Ala511.

The albumin portion of an albumin fusion protein of the invention may comprise at least one subdomain or domain of HA or conservative modifications thereof. If the fusion is based on subdomains, some or all of the adjacent linker may optionally be used to link to the Therapeutic protein moiety.

### **Albumin Fusion Proteins**

The present invention relates generally to albumin fusion proteins and methods of treating, preventing, or ameliorating diseases or disorders. As used herein, "albumin fusion

protein” refers to a protein formed by the fusion of at least one molecule of albumin (or a fragment or variant thereof) to at least one molecule of a Therapeutic protein (or fragment or variant thereof). An albumin fusion protein of the invention comprises at least a fragment or variant of a Therapeutic protein and at least a fragment or variant of human serum albumin, which are associated with one another, such as by genetic fusion (i.e., the albumin fusion protein is generated by translation of a nucleic acid in which a polynucleotide encoding all or a portion of a Therapeutic protein is joined in-frame with a polynucleotide encoding all or a portion of albumin) to one another. The Therapeutic protein and albumin protein, once part of the albumin fusion protein, may be referred to as a “portion”, “region”, or “moiety” of the albumin fusion protein.

In one embodiment, the invention provides an albumin fusion protein comprising, or alternatively consisting of, a Therapeutic protein and a serum albumin protein. In other embodiments, the invention provides an albumin fusion protein comprising, or alternatively consisting of, a biologically active and/or therapeutically active fragment of a Therapeutic protein and a serum albumin protein. In other embodiments, the invention provides an albumin fusion protein comprising, or alternatively consisting of, a biologically active and/or therapeutically active variant of a Therapeutic protein and a serum albumin protein. In some embodiments, the serum albumin protein component of the albumin fusion protein is the mature portion of serum albumin.

In further embodiments, the invention provides an albumin fusion protein comprising, or alternatively consisting of, a Therapeutic protein, and a biologically active and/or therapeutically active fragment of serum albumin. In further embodiments, the invention provides an albumin fusion protein comprising, or alternatively consisting of, a Therapeutic protein and a biologically active and/or therapeutically active variant of serum albumin. In certain embodiments, the Therapeutic protein portion of the albumin fusion protein is the mature portion of the Therapeutic protein.

In further embodiments, the invention provides an albumin fusion protein comprising, or alternatively consisting of, a biologically active and/or therapeutically active fragment or variant of a Therapeutic protein and a biologically active and/or therapeutically active fragment or variant of serum albumin. In some embodiments, the invention provides an albumin fusion protein comprising, or alternatively consisting of, the mature portion of a Therapeutic protein and the mature portion of serum albumin.

The albumin fusion protein comprises HA as the N-terminal portion, and a Therapeutic protein as the C-terminal portion. Alternatively, an albumin fusion protein



- 15 -

comprising HA as the C-terminal portion, and a Therapeutic protein as the N-terminal portion may also be used.

In other embodiments, the albumin fusion protein has a Therapeutic protein fused to both the N-terminus and the C-terminus of albumin. In one embodiment, the Therapeutic proteins fused at the N- and C- termini are the same Therapeutic proteins. In another embodiment, the Therapeutic proteins fused at the N- and C- termini are different Therapeutic proteins. In yet another embodiment, the Therapeutic proteins fused at the N- and C- termini are different Therapeutic proteins which may be used to treat or prevent the same disease, disorder, or condition. In some embodiments, the Therapeutic proteins fused at the N- and C- termini are different Therapeutic proteins which may be used to treat or prevent diseases or disorders which are known in the art to commonly occur in patients simultaneously.

In addition to albumin fusion protein in which the albumin portion is fused N- terminal and/or C-terminal of the Therapeutic protein portion, albumin fusion proteins of the invention may also be produced by inserting the Therapeutic protein or peptide of interest into an internal region of HA. For instance, within the protein sequence of the HA molecule a number of loops or turns exist between the end and beginning of  $\alpha$ -helices, which are stabilized by disulphide bonds. The loops, as determined from the crystal structure of HA (PDB identifiers 1AO6, 1BJ5, 1BKE, 1BM0, 1E7E to 1E7I and 1UOR) for the most part extend away from the body of the molecule. These loops are useful for the insertion, or internal fusion, of therapeutically active peptides, particularly those requiring a secondary structure to be functional, or Therapeutic proteins, to essentially generate an albumin molecule with specific biological activity.

Loops in human albumin structure into which peptides or polypeptides may be inserted to generate albumin fusion proteins of the invention include: Val54-Asn61, Thr76-Asp89, Ala92-Glu100, Gln170-Ala176, His247-Glu252, Glu266-Glu277, Glu280-His288, Ala362-Glu368, Lys439-Pro447, Val462-Lys475, Thr478-Pro486, and Lys560-Thr566. In other embodiments, peptides or polypeptides are inserted into the Val54-Asn61, Gln170-Ala176, and/or Lys560-Thr566 loops of mature human albumin (Table 1) (SEQ ID NO:18).

The Therapeutic protein to be inserted may be derived from any source, including phage display and synthetic peptide libraries screened for specific biological activity or from the active portions of a molecule with the desired function. Additionally, random peptide libraries comprising Kunitz domain peptides that are candidates for use as a Therapeutic protein may be generated within particular loops or by insertions of such randomized peptides

- 16 -

into particular loops of the HA molecule and in which many (e.g.  $5 \times 10^9$ ) combinations of amino acids are represented.

Such library(s) could be generated on HA or domain fragments of HA by one of the following methods:

(a) randomized mutation of amino acids within one or more peptide loops of HA or HA domain fragments. Either one, more than one or all the residues within a loop could be mutated in this manner;

(b) replacement of, or insertion into one or more loops of HA or HA domain fragments (*i.e.*, internal fusion) of a randomized peptide(s) of length  $X_n$  (where  $X$  is an amino acid and  $n$  is the number of residues;

(c) N-, C- or N- and C- terminal peptide/protein fusions in addition to (a) and/or (b).

The HA or HA domain fragment may also be made multifunctional by grafting the peptides derived from different screens of different loops against different targets into the same HA or HA domain fragment.

Non-limiting examples of peptides inserted into a loop of human serum albumin are DX-890 (an inhibitor of human neutrophil elastase), DPI-14 (an inhibitor of human neutrophil elastase), DX-88 peptide (an inhibitor of human plasma kallikrein, Table 2), and DX-1000 (an inhibitor of human plasmin, Table 2) or peptide fragments or peptide variants thereof. More particularly, the invention encompasses albumin fusion proteins which comprise peptide fragments or peptide variants at least 7 at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 20, at least 25, at least 30, at least 35, or at least 40 amino acids in length inserted into a loop of human serum albumin. The invention also encompasses albumin fusion proteins which comprise peptide fragments or peptide variants at least 7 at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 20, at least 25, at least 30, at least 35, or at least 40 amino acids fused to the N-terminus of human serum albumin. The invention also encompasses albumin fusion proteins which comprise peptide fragments or peptide variants at least 7 at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 20, at least 25, at least 30, at least 35, or at least 40 amino acids fused to the C-terminus of human serum albumin.

Generally, the albumin fusion proteins of the invention may have one HA-derived region and one Therapeutic protein-derived region. Multiple regions of each protein, however, may be used to make an albumin fusion protein of the invention. Similarly, more than one Therapeutic protein may be used to make an albumin fusion protein of the invention.

- 17 -

For instance, a Therapeutic protein may be fused to both the N- and C-terminal ends of the HA. In such a configuration, the Therapeutic protein portions may be the same or different Therapeutic protein molecules. The structure of bifunctional albumin fusion proteins may be represented as: X-HA-Y or Y-HA-X or X-Y-HA or HA-X-Y or HA-X-Y-HA or HA-Y-X-HA or HA-X-X-HA or HA-Y-Y-HA or HA-X-HA-Y or X-HA-Y-HA or multiple combinations or inserting X and/or Y within the HA sequence at any location.

Additional embodiments that involve a therapeutic protein "X", such as a Kunitz domain, and a therapeutic peptide "Y" involve separating HA into parts 1 and 2. The fusion proteins of the invention could have the forms: X-HA(part1)-Y-HA(part2) and HA(part1)-Y-HA(part2)-X. Additional embodiments involve two therapeutic protein domains "X" and "Z" and a therapeutic peptide "Y" leading to fusion proteins of the forms: X-HA(part1)-Y-HA(part2)-Z and Z-HA(part1)-Y-HA(part2)-X.

Bi- or multi-functional albumin fusion proteins may be prepared in various ratios depending on function, half-life, etc.

Bi- or multi-functional albumin fusion proteins may also be prepared to target the Therapeutic protein portion of a fusion to a target organ or cell type via protein or peptide at the opposite terminus of HA.

As an alternative to the fusion of known therapeutic molecules, the peptides could be obtained by screening libraries constructed as fusions to the N-, C- or N- and C- termini of HA, or domain fragment of HA, of typically 6, 8, 12, 20 or 25 or  $X_n$  (where X is an amino acid (aa) and n equals the number of residues) randomized amino acids, and in which all possible combinations of amino acids were allowed. A particular advantage of this approach is that the peptides may be selected *in situ* on the HA molecule and the properties of the peptide would therefore be as selected for rather than, potentially, modified as might be the case for a peptide derived by any other method then being attached to HA. Such selection is not needed for attachment of well-folded domains, such as Kunitz domains, at the ends of HA. Selection *in-situ* is likely to be important for peptides that have no disulfides or a single disulfide loop.

Additionally, the albumin fusion proteins of the invention may include a linker peptide between the fused portions to provide greater physical separation between the moieties and thus maximize the accessibility of the Therapeutic protein portion, for instance, for binding to its cognate receptor. The linker peptide may consist of amino acids such that it is flexible or more rigid.

- 18 -

Therefore, as described above, the albumin fusion proteins of the invention may have the following formula R2-R1; R1-R2; R2-R1-R2; R2-L-R1-L-R2; R1-L-R2; R2-L-R1; or R1-L-R2-L-R1, wherein R1 is at least one Therapeutic protein, peptide or polypeptide sequence (including fragments or variants thereof), and not necessarily the same Therapeutic protein, L is a linker and R2 is a serum albumin sequence (including fragments or variants thereof). Exemplary linkers include (GGGGS)<sub>N</sub> (SEQ ID NO:\_\_\_\_) or (GGGS)<sub>N</sub> (SEQ ID NO:\_\_\_\_) or (GGS)<sub>N</sub>, wherein N is an integer greater than or equal to 1 and wherein G represents glycine and S represents serine.

In certain embodiments, albumin fusion proteins of the invention comprising a Therapeutic protein have extended shelf life or *in vivo* half-life or therapeutic activity compared to the shelf life or *in vivo* half-life or therapeutic activity of the same Therapeutic protein when not fused to albumin. Shelf-life typically refers to the time period over which the therapeutic activity of a Therapeutic protein in solution or in some other storage formulation, is stable without undue loss of therapeutic activity. Many of the Therapeutic proteins are highly labile in their unfused state. As described below, the typical shelf-life of these Therapeutic proteins is markedly prolonged upon incorporation into the albumin fusion protein of the invention.

Albumin fusion proteins of the invention with "prolonged" or "extended" shelf-life exhibit greater therapeutic activity relative to a standard that has been subjected to the same storage and handling conditions. The standard may be the unfused full-length Therapeutic protein. When the Therapeutic protein portion of the albumin fusion protein is an analog, a variant, or is otherwise altered or does not include the complete sequence for that protein, the prolongation of therapeutic activity may alternatively be compared to the unfused equivalent of that analog, variant, altered peptide or incomplete sequence. As an example, an albumin fusion protein of the invention may retain greater than about 100% of the therapeutic activity, or greater than about 105%, 110%, 120%, 130%, 150% or 200% of the therapeutic activity of a standard when subjected to the same storage and handling conditions as the standard when compared at a given time point. However, it is noted that the therapeutic activity depends on the Therapeutic protein's stability, and may be below 100%.

Shelf-life may also be assessed in terms of therapeutic activity remaining after storage, normalized to therapeutic activity when storage began. Albumin fusion proteins of the invention with prolonged or extended shelf-life as exhibited by prolonged or extended therapeutic activity may retain greater than about 50% of the therapeutic activity, about 60%,

- 19 -

70%, 80%, or 90% or more of the therapeutic activity of the equivalent unfused Therapeutic protein when subjected to the same conditions.

Albumin fusion proteins of the invention exhibit greater solubility relative to the non-fused Therapeutic protein standard that has been subjected to the same storage and handling conditions.

### **Therapeutic proteins**

As stated above, an albumin fusion protein of the invention comprises at least a fragment or variant of a Therapeutic protein and at least a fragment or variant of human serum albumin, which are associated with one another by genetic fusion.

As used herein, "Therapeutic protein" refers to a Kunitz domain peptide, non-limiting examples of which include DX-890, DPI-14, DX-88 or DX-1000, or fragments or variants thereof, having one or more therapeutic and/or biological activities. A Kunitz domain is a folding domain of approximately 51-64 residues which forms a central anti-parallel beta sheet and a short C-terminal helix. This characteristic domain comprises six cysteine residues that form three disulfide bonds, resulting in a double-loop structure. Between the N-terminal region and the first beta strand resides the active inhibitory binding loop. This binding loop is disulfide bonded through the P2 C<sub>14</sub> residue to the hairpin loop formed between the last two beta strands.

A Kunitz domain is a polypeptide of from about 51 AAs to about 64 AAs of the form:

X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>X<sub>4</sub>C<sub>5</sub>X<sub>6</sub>X<sub>7</sub>X<sub>8</sub>X<sub>9</sub>X<sub>9a</sub>X<sub>10</sub>X<sub>11</sub>X<sub>12</sub>X<sub>13</sub>C<sub>14</sub>X<sub>15</sub>X<sub>16</sub>X<sub>17</sub>X<sub>18</sub>X<sub>19</sub>X<sub>20</sub>X<sub>21</sub>X<sub>22</sub>X<sub>23</sub>X<sub>24</sub>X<sub>25</sub>X<sub>26</sub>X<sub>26a</sub>X<sub>26b</sub>-  
 X<sub>26c</sub>X<sub>27</sub>X<sub>28</sub>X<sub>29</sub>C<sub>30</sub>X<sub>31</sub>X<sub>32</sub>X<sub>33</sub>X<sub>34</sub>X<sub>35</sub>X<sub>36</sub>X<sub>37</sub>C<sub>38</sub>X<sub>39</sub>X<sub>40</sub>X<sub>41</sub>X<sub>42</sub>X<sub>42a</sub>X<sub>42b</sub>X<sub>43</sub>X<sub>44</sub>X<sub>45</sub>X<sub>46</sub>X<sub>47</sub>X<sub>48</sub>X<sub>49</sub>-  
 C<sub>50</sub>X<sub>51</sub>X<sub>52</sub>X<sub>53</sub>X<sub>54</sub>C<sub>55</sub>X<sub>56</sub>X<sub>57</sub>X<sub>58</sub> (SEQ ID NO: \_\_\_\_\_)

Disulfides are formed between C<sub>5</sub> and C<sub>55</sub>, C<sub>14</sub> and C<sub>38</sub>, and C<sub>30</sub> and C<sub>51</sub>. The C<sub>14</sub>-C<sub>38</sub> disulfide is always seen in natural Kunitz domains, but may be removed in artificial Kunitz domains. If C<sub>14</sub> is changed to another amino-acid type, then C<sub>38</sub> is also changed to a non-cysteine and *vice versa*. Any polypeptide may be fused to the amino terminus. X<sub>1</sub>-X<sub>4</sub> may comprise zero to four amino acids. X<sub>6</sub>-X<sub>13</sub> may comprise 8 or 9 amino acids. If X<sub>9a</sub> is absent, then X<sub>12</sub> is Gly. Each of X<sub>26a</sub>, X<sub>26b</sub>, and X<sub>26c</sub> may be absent; that is, X<sub>15</sub>-X<sub>30</sub> may comprise 16, 17, 18, or 19 amino acids. X<sub>33</sub> is Phe or Tyr. X<sub>39</sub>-X<sub>50</sub> may comprise 12, 13, 14, or 15 amino acids; that is, each of X<sub>42a</sub>, X<sub>42b</sub>, and X<sub>42c</sub> may be absent. X<sub>45</sub> is Phe or Tyr. X<sub>56</sub>-X<sub>58</sub> may comprise zero to three amino acids. Additional cysteines may occur at positions 50, 53, 54 or

58. Any polypeptide may be fused to the carboxy terminus. Table 3 shows the amino-acid sequences of 21 known human Kunitz domains.

**Table 3:** Amino acid sequences of 21 known human Kunitz domains

Domain	Protein	Amino Acid Sequence	Accession
single	A4 (amyloid precursor PTN)	VREVCSEQAETGPCRAMISRWFYFDVTEGK CAPFFYGGCGGNRRNFDTEEYCMVCGSA SEQ ID NO: ____	SP:A4_HUMAN A# P05067
single	embl loCus HS461P17 = "CAB37"	KQDVCEMPKETGPCLAYFLHWYDKKDNT CSMFVYGGCQGNRRNFQSKANCLNTCKNK SEQ ID NO: ____	(CAB37635; g4467797)
single	Amyloid-like PTN 2	VKAVCSQEAMTGPCRAVMPRWYFDLSKGK CVRFIYGGCGGNRRNFESDYCMVCKAM SEQ ID NO: ____	Loc:1703344;S41082 g1082207 & g1703344 & g477608
K1	ITI	KEDSCQLGYSAGPCMGMTSRYFYNGTSMA CETFQYGGCMGNGNRFVTEKECLQTCRTV SEQ ID NO: ____	SP:HC_HUMAN, A# P02760 (HI-8e) = gi 223133
K2	ITI	TVAACNLPIVRGPCRAFIQLWAFDAVKGK CVLFPHYGGCQGNNGKFYSEKECREYCGVP SEQ ID NO: ____	SP:HC_HUMAN, A# P02760 (HI-8e) = gi 223133
K1	TFPI-1 = LACI	MHSFCAFKADDGPCKAIMKRFFFNIFTRQ CEEFIYGGCEGNQNRFFESLEECKKMCTRD N SEQ ID NO: ____ (corrected 2000.05.14)	SP:LACI_HUMAN, A# P10646 gim 14667
K2	TFPI-1	KPDFCFLEEDPGICRGYITRYFYNNQTKQ CERFKYGGCLGNMNNFETLEECKNICEDG SEQ ID NO: ____	SP:LACI_HUMAN, A# P10646 gim 14667
K3	TFPI-1	GPSWCLTPADRGLCRANENRFYNSVIGK CRPFKYSGCGGNENNFTSKQECLRACKKG SEQ ID NO: ____	SP:LACI_HUMAN, A# P10646 gim 14667
K1	TFPI-2	NAEICLLPLDYGPCRALLRYYYDRYTQS CRQFLYGGCEGNANNFYTWACDDACWRI SEQ ID NO: ____	Specher &al. PNAS 91:3353-3357 (1994)
K2	TFPI-2	VPKVCRLQVDDQCEGSTKEYFFNLSSMT CEKFFSGGCHRRNRFPDEATCMGFCAPK SEQ ID NO: ____	Specher &al, PNAS 91:3353ff (1994)
K3	TFPI-2	IPSFYCSPKDEGLCSANVTRYFYFNPRYRT CDAFTYTGCGGNDNNFVSREDCKRACAKA SEQ ID NO: ____	Specher &al, PNAS 91:3353ff (1994)

- 21 -

Domain	Protein	Amino Acid Sequence	Accession
K1	Hepatocyte GF activator inhib type 1	TEDYCLASNKVGRCRGSFPRWYYDPTEQI CKSFVYGGCLGNKNNYLREEECILACRGV SEQ ID NO: ____	Locus 2924601
K2	Hepatocyte GF activator inhib type 1	DKGHCVLDLPDTGLCKESI PRWYYNPFSEH CARFTYGGCYGNKNNFEEEQQCLES CRGI SEQ ID NO: ____	Locus 2924601
K1	hepatocyte GF activator inhib. type 2	IHDFCLVSKVVGRCRASMPRWYNTDGS CQLFVYGGCDGNSNNYLTKEECLKKCATV SEQ ID NO: ____	LOC. 2924620
K2	hepatocyte GF activator inhib. type2	YEEYCTANAVTGPCRASFPRWYFDVERNS CNNFIYGGCRGNKNSYRSEEACMLRCFRQ SEQ ID NO: ____	LOC. 2924620
Single	PRF	TVAACNLPVIRGPCRAFIQLWAFDAVKGK CVLFPYGGCQGNNGKFYSEKECREYCGVP SEQ ID NO: ____	gi 223132 Name: 0511271A
Single	HKI B9 domain	LPNVCAFPMEKGPCQTYMTRWFFNFETGE CELFA YGGCGGNSNNFLRKEKCEKFKFT SEQ ID NO: ____	gi 579567 WO93/14123-A; g542925
Single	Collagen V1 (VII)	SDDPCSLPLDEGSCTAYTLRWYHRAVTEA CHPFVYGGCGGNANRFGTREACERRCPPR SEQ ID NO: ____	NCBI: gi 543915
Single	collagen alpha 1(VII)	EDDPCSLPLDEGSCTAYTLRWYHRAVTGS TEACHPFVYGGCGGNANRFGTREACERRC PPR SEQ ID NO: ____	g627406 - A54849 GI:627406
Single	collagen V3	ETDICKLPKDEGTCRDFILKWYDPNTKS CARFWYGGCGGNENKFGSQKECEKVCAPV SEQ ID NO: ____	NCBI Seq ID: 512802 WO93/14119-A. 2193976 (Xray)
single	Chromosome 20 ptn "Chrome20"	FQEPCLMPVRHGNCNHEAQRWHFDFKNYR CTPFKYRGCEGNANNFLNEDACRTACMLI SEQ ID NO: ____	CAB37634 PID g7024350

Any of the domains in Table 1 could be engineered to have a specific biological effect (such as inhibiting a particular protease) and be fused to HA. Thus an albumin fusion protein of the invention may contain at least a fragment or variant of a Therapeutic protein. Variants include mutants, analogs, and mimetics, as well as homologs, including the endogenous or naturally occurring correlates.

- 22 -

By a polypeptide displaying a “therapeutic activity” or a protein that is “therapeutically active” is meant a polypeptide that possesses one or more known biological and/or therapeutic activities associated with a Therapeutic protein such as one or more of the Therapeutic proteins described herein or otherwise known in the art. As a non-limiting example, a “Therapeutic protein” is a protein that is useful to treat, prevent or ameliorate a disease, condition or disorder.

As used herein, “therapeutic activity” or “activity” may refer to an activity whose effect is consistent with a desirable therapeutic outcome in humans, or to desired effects in non-human mammals or in other species or organisms. Therapeutic activity may be measured *in vivo* or *in vitro*. For example, a desirable effect may be assayed in cell culture. Such *in vitro* or cell culture assays are commonly available for many Therapeutic proteins as described in the art.

Examples of useful assays include, but are not limited to, those described in references and publications of Table 4, specifically incorporated by reference herein, and those described in the Examples herein. The activity exhibited by the fusion proteins of the invention may be measured, for example, by easily performed *in vitro* assays, such as those described herein. Using these assays, such parameters as the relative biological and/or therapeutic activity that the fusion proteins exhibit as compared to the Therapeutic protein (or fragment or variant thereof) when it is not fused to albumin can be determined.

Therapeutic proteins corresponding to a Therapeutic protein portion of an albumin fusion protein of the invention may be modified by the attachment of one or more oligosaccharide groups. The modification, referred to as glycosylation, can dramatically affect the physical properties of proteins and can be important in protein stability, secretion, and localization. Such modifications are described in detail in U.S. Provisional Application Serial No. 60/355,547 and WO 01/79480, which are incorporated herein by reference.

Therapeutic proteins corresponding to a Therapeutic protein portion of an albumin fusion protein of the invention, as well as analogs and variants thereof, may be modified so that glycosylation at one or more sites is altered as a result of manipulation(s) of their nucleic acid sequence, by the host cell in which they are expressed, or due to other conditions of their expression. For example, glycosylation isomers may be produced by abolishing or introducing glycosylation sites, *e.g.*, by substitution or deletion of amino acid residues, such as substitution of glutamine for asparagine, or unglycosylated recombinant proteins may be produced by expressing the proteins in host cells that will not glycosylate them, *e.g.* in *E. coli* or glycosylation-deficient yeast. Examples of these approaches are described in more detail in



- 23 -

U.S. Provisional Application Serial No. 60/355,547 and WO 01/79480, which are incorporated by reference, and are known in the art.

Table 4 provides a non-exhaustive list of Therapeutic proteins that correspond to a Therapeutic protein portion of an albumin fusion protein of the invention. The "Therapeutic Protein X" column discloses Therapeutic protein molecules followed by parentheses containing scientific and brand names that comprise, or alternatively consist of, that Therapeutic protein molecule or a fragment or variant thereof. "Therapeutic protein X" as used herein may refer either to an individual Therapeutic protein molecule (as defined by the amino acid sequence obtainable from the CAS and Genbank accession numbers), or to the entire group of Therapeutic proteins associated with a given Therapeutic protein molecule disclosed in this column. The information associated with each of these entries are each incorporated by reference in their entirety, particularly with respect to the amino acid sequences described therein. The "PCT/Patent Reference" column provides U.S. Patent numbers, or PCT International Publication Numbers corresponding to patents and/or published patent applications that describe the Therapeutic protein molecule. Each of the patents and/or published patent applications cited in the "PCT/Patent Reference" column are herein incorporated by reference in their entirety. In particular, the amino acid sequences of the specified polypeptide set forth in the sequence listing of each cited "PCT/Patent Reference", the variants of these amino acid sequences (mutations, fragments, etc.) set forth, for example, in the detailed description of each cited "PCT/Patent Reference", the therapeutic indications set forth, for example, in the detailed description of each cited "PCT/Patent Reference", and the activity assays for the specified polypeptide set forth in the detailed description, and more particularly, the examples of each cited "PCT/Patent Reference" are incorporated herein by reference. The "Biological activity" column describes Biological activities associated with the Therapeutic protein molecule. Each of the references cited in the "Relevant Publications" column are herein incorporated by reference in their entirety, particularly with respect to the description of the respective activity assay described in the reference (see Methods section, for example) for assaying the corresponding biological activity. The "Preferred Indication Y" column describes disease, disorders, and/or conditions that may be treated, prevented, diagnosed, or ameliorated by Therapeutic protein X or an albumin fusion protein of the invention comprising a Therapeutic protein X portion.

**Table 4: A List of Selected Therapeutic Proteins**

Therapeutic Protein X	PCT/Patent Reference	Biological Activity	Relevant Publications	Preferred Indication Y
DX-890, DPI14	U.S. Patent No. 5,663,143, SEQ ID NO:20 = DX-890	Inhibition of human neutrophil elastase, $K_i \sim 5$ pM.	Rusckowski et al. (2000) J. Nuclear Medicine 41:363-74	Emphysema, Cystic fibrosis COPD, Bronchitis, Pulmonary Hypertension, Acute respiratory distress syndrome, Interstitial lung disease, Asthma, Smoke intoxication, Bronchopulmonary dysplasia, Pneumonia, Thermal Injury, Lung transplant rejection.
DX-88	U.S. Patent Nos. 6,333,402; 5,994,125; 6,057,287; and 5,795,865	Inhibition of human plasma kallikrein	Markland et al. <i>Biochemistry</i> 35(24):8058-67, 1996. Ley et al. (1996) <i>Mol Divers</i> 2(1-2)119-24.	HAE
DX-1000	U.S. Patent Nos. 6,010,880; 6,071,723; and 6,103,499	Inhibits human plasmin	Markland et al. <i>Biochemistry</i> 35(24):8045-57, 1996. Ley et al. (1996) <i>Mol Divers</i> 2(1-2)119-24.	Bleeding, cancer.

In various embodiments, the albumin fusion proteins of the invention are capable of a therapeutic activity and/or biologic activity corresponding to the therapeutic activity and/or biologic activity of the Therapeutic protein corresponding to the Therapeutic protein portion of the albumin fusion protein listed in the corresponding row of Table 4. (See, e.g., the “Biological Activity” and “Therapeutic Protein X” columns of Table 4.) In other embodiments, the therapeutically active protein portions of the albumin fusion proteins of the invention are fragments or variants of the reference sequence and are capable of the therapeutic activity and/or biologic activity of the corresponding Therapeutic protein disclosed in “Biological Activity” column of Table 4.

**Polypeptide and Polynucleotide Fragments and Variants***Fragments*

The present invention is further directed to fragments of the Therapeutic proteins described in Table 4, albumin proteins, and/or albumin fusion proteins of the invention.

Even if deletion of one or more amino acids from the N-terminus of a protein results in modification or loss of one or more biological functions of the Therapeutic protein, albumin protein, and/or albumin fusion protein, other Therapeutic activities and/or functional activities (e.g., biological activities, ability to multimerize, ability to bind a ligand) may still be retained. For example, the ability of polypeptides with N-terminal deletions to induce and/or bind to antibodies which recognize the complete or mature forms of the polypeptides generally will be retained when less than the majority of the residues of the complete polypeptide are removed from the N-terminus. Whether a particular polypeptide lacking N-terminal residues of a complete polypeptide retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a mutein with a large number of deleted N-terminal amino acid residues may retain some biological or immunogenic activities. In fact, peptides composed of as few as six amino acid residues may often evoke an immune response.

Accordingly, fragments of a Therapeutic protein corresponding to a Therapeutic protein portion of an albumin fusion protein of the invention, include the full length protein as well as polypeptides having one or more residues deleted from the amino terminus of the amino acid sequence of the reference polypeptide (e.g., a Therapeutic protein as disclosed in Table 4). Polynucleotides encoding these polypeptides are also encompassed by the invention.

In addition, fragments of serum albumin polypeptides corresponding to an albumin protein portion of an albumin fusion protein of the invention, include the full length protein as well as polypeptides having one or more residues deleted from the amino terminus of the amino acid sequence of the reference polypeptide (i.e., serum albumin). Polynucleotides encoding these polypeptides are also encompassed by the invention.

Moreover, fragments of albumin fusion proteins of the invention include the full-length albumin fusion protein as well as polypeptides having one or more residues deleted from the amino terminus of the albumin fusion protein. Polynucleotides encoding these polypeptides are also encompassed by the invention.

Also as mentioned above, even if deletion of one or more amino acids from the N-terminus or C-terminus of a reference polypeptide (e.g., a Therapeutic protein and/or serum

- 26 -

albumin protein) results in modification or loss of one or more biological functions of the protein, other functional activities (e.g., biological activities, ability to multimerize, ability to bind a ligand) and/or Therapeutic activities may still be retained. For example the ability of polypeptides with C-terminal deletions to induce and/or bind to antibodies which recognize the complete or mature forms of the polypeptide generally will be retained when less than the majority of the residues of the complete or mature polypeptide are removed from the C-terminus. Whether a particular polypeptide lacking the N-terminal and/or C-terminal residues of a reference polypeptide retains Therapeutic activity can readily be determined by routine methods described herein and/or otherwise known in the art.

The present invention further provides polypeptides having one or more residues deleted from the carboxy terminus of the amino acid sequence of a Therapeutic protein corresponding to a Therapeutic protein portion of an albumin fusion protein of the invention (e.g., a Therapeutic protein referred to in Table 4). Polynucleotides encoding these polypeptides are also encompassed by the invention.

In addition, the present invention provides polypeptides having one or more residues deleted from the carboxy terminus of the amino acid sequence of an albumin protein corresponding to an albumin protein portion of an albumin fusion protein of the invention (e.g., serum albumin). Polynucleotides encoding these polypeptides are also encompassed by the invention.

Moreover, the present invention provides polypeptides having one or more residues deleted from the carboxy terminus of an albumin fusion protein of the invention. Polynucleotides encoding these polypeptides are also encompassed by the invention.

In addition, any of the above described N- or C-terminal deletions can be combined to produce a N- and C-terminal deleted reference polypeptide (e.g., a Therapeutic protein referred to in Table 4, or serum albumin (e.g., SEQ ID NO:18, Table 1), or an albumin fusion protein of the invention). The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini. Polynucleotides encoding these polypeptides are also encompassed by the invention.

The present application is also directed to proteins containing polypeptides at least 60%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a reference polypeptide sequence (e.g., a Therapeutic protein, serum albumin protein or an albumin fusion protein of the invention) set forth herein, or fragments thereof. In some embodiments, the application is directed to proteins comprising polypeptides at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to reference polypeptides having the amino acid sequence of N- and

- 27 -

C-terminal deletions as described above. Polynucleotides encoding these polypeptides are also encompassed by the invention.

Other polypeptide fragments of the invention are fragments comprising, or alternatively, consisting of, an amino acid sequence that displays a Therapeutic activity and/or functional activity (e.g. biological activity) of the polypeptide sequence of the Therapeutic protein or serum albumin protein of which the amino acid sequence is a fragment.

Other polypeptide fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

#### *Variants*

“Variant” refers to a polynucleotide or nucleic acid differing from a reference nucleic acid or polypeptide, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the reference nucleic acid or polypeptide.

As used herein, “variant”, refers to a Therapeutic protein portion of an albumin fusion protein of the invention, albumin portion of an albumin fusion protein of the invention, or albumin fusion protein differing in sequence from a Therapeutic protein (e.g., see “Therapeutic Protein X” column of Table 4), albumin protein, and/or albumin fusion protein of the invention, respectively, but retaining at least one functional and/or therapeutic property thereof (e.g., a therapeutic activity and/or biological activity as disclosed in the “Biological Activity” column of Table 4) as described elsewhere herein or otherwise known in the art. Generally, variants are overall very similar, and, in many regions, identical to the amino acid sequence of the Therapeutic protein corresponding to a Therapeutic protein portion of an albumin fusion protein of the invention, albumin protein corresponding to an albumin protein portion of an albumin fusion protein of the invention, and/or albumin fusion protein of the invention. Nucleic acids encoding these variants are also encompassed by the invention.

The present invention is also directed to proteins which comprise, or alternatively consist of, an amino acid sequence which is at least 60%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100%, identical to, for example, the amino acid sequence of a Therapeutic protein corresponding to a Therapeutic protein portion of an albumin fusion protein of the invention (e.g., an amino acid sequence disclosed in a reference in Table 4, or fragments or variants thereof), albumin proteins (e.g., Table 1) or fragments or variants thereof corresponding to an albumin protein portion of an albumin fusion protein of the invention, and/or albumin fusion proteins of the invention. Fragments of these polypeptides are also

- 28 -

provided (e.g., those fragments described herein). Further polypeptides encompassed by the invention are polypeptides encoded by polynucleotides which hybridize to the complement of a nucleic acid molecule encoding an amino acid sequence of the invention under stringent hybridization conditions (e.g., hybridization to filter bound DNA in 6X Sodium chloride/Sodium citrate (SSC) at about 45 degrees Celsius, followed by one or more washes in 0.2X SSC, 0.1% SDS at about 50 - 65 degrees Celsius), under highly stringent conditions (e.g., hybridization to filter bound DNA in 6X sodium chloride/Sodium citrate (SSC) at about 45 degrees Celsius, followed by one or more washes in 0.1X SSC, 0.2% SDS at about 68 degrees Celsius), or under other stringent hybridization conditions which are known to those of skill in the art (see, for example, Ausubel, F.M. et al., eds., 1989 *Current protocol in Molecular Biology*, Green publishing associates, Inc., and John Wiley & Sons Inc., New York, at pages 6.3.1 - 6.3.6 and 2.10.3). Polynucleotides encoding these polypeptides are also encompassed by the invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, or substituted with another amino acid. These alterations of the reference sequence may occur at the amino- or carboxy-terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 60%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequence of an albumin fusion protein of the invention or a fragment thereof (such as the Therapeutic protein portion of the albumin fusion protein or the albumin portion of the albumin fusion protein), can be determined conventionally using known computer programs. Such programs and methods of using them are described, e.g., in U.S. Provisional Application Ser. No. 60/355,547 and WO 01/79480 (pp. 41-43), which are incorporated by reference herein, and are well known in the art.

The polynucleotide variants of the invention may contain alterations in the coding regions, non-coding regions, or both. Polynucleotide variants include those containing

- 29 -

alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Such nucleotide variants may be produced by silent substitutions due to the degeneracy of the genetic code. Polypeptide variants include those in which less than 50, less than 40, less than 30, less than 20, less than 10, or 5-50, 5-25, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a microbial host, such as, yeast or *E. coli*).

In another embodiment, a polynucleotide encoding an albumin portion of an albumin fusion protein of the invention is optimized for expression in yeast or mammalian cells. In yet another embodiment, a polynucleotide encoding a Therapeutic protein portion of an albumin fusion protein of the invention is optimized for expression in yeast or mammalian cells. In still another embodiment, a polynucleotide encoding an albumin fusion protein of the invention is optimized for expression in yeast or mammalian cells.

In an alternative embodiment, a codon optimized polynucleotide encoding a Therapeutic protein portion of an albumin fusion protein of the invention does not hybridize to the wild type polynucleotide encoding the Therapeutic protein under stringent hybridization conditions as described herein. In a further embodiment, a codon optimized polynucleotide encoding an albumin portion of an albumin fusion protein of the invention does not hybridize to the wild type polynucleotide encoding the albumin protein under stringent hybridization conditions as described herein. In another embodiment, a codon optimized polynucleotide encoding an albumin fusion protein of the invention does not hybridize to the wild type polynucleotide encoding the Therapeutic protein portion or the albumin protein portion under stringent hybridization conditions as described herein.

In an additional embodiment, polynucleotides encoding a Therapeutic protein portion of an albumin fusion protein of the invention do not comprise, or alternatively consist of, the naturally occurring sequence of that Therapeutic protein. In a further embodiment, polynucleotides encoding an albumin protein portion of an albumin fusion protein of the invention do not comprise, or alternatively consist of, the naturally occurring sequence of albumin protein. In an alternative embodiment, polynucleotides encoding an albumin fusion protein of the invention do not comprise, or alternatively consist of, the naturally occurring sequence of a Therapeutic protein portion or the albumin protein portion.

- 30 -

In an additional embodiment, the Therapeutic protein may be selected from a random peptide library by biopanning, as there will be no naturally occurring wild type polynucleotide.

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985)). These allelic variants can vary at either the polynucleotide and/or polypeptide level and are included in the present invention. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids may be deleted from the N-terminus or C-terminus of the polypeptide of the present invention without substantial loss of biological function. See, e.g., Ron et al. (J. Biol. Chem. 268: 2984-2988 (1993) (KGF variants) and Dobeli et al., J. Biotechnology 7:199-216 (1988) (interferon gamma variants).

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein (e.g. Gayle and coworkers (J. Biol. Chem. 268:22105-22111 (1993) (IL-1a variants))). Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which have a functional activity (e.g., biological activity and/or therapeutic activity). In further embodiments the invention provides variants of albumin fusion proteins that have a functional activity (e.g., biological activity and/or therapeutic activity, such as that disclosed in the "Biological Activity" column in Table 4) that corresponds to one or more biological and/or therapeutic activities of the Therapeutic protein corresponding to the Therapeutic protein portion of the albumin fusion protein. Such variants include deletions, insertions, inversions, repeats, and



- 31 -

substitutions selected according to general rules known in the art so as have little effect on activity.

In other embodiments, the variants of the invention have conservative substitutions. By "conservative substitutions" is intended swaps within groups such as replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Guidance concerning how to make phenotypically silent amino acid substitutions is provided, for example, in Bowie et al., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

As the authors state, proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly. Besides conservative amino acid substitution, variants of the present invention include (i) polypeptides containing substitutions of one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) polypeptides containing substitutions of one or more of the amino acid residues having a substituent group, or (iii) polypeptides which have been fused with or chemically conjugated to another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), (iv) polypeptide containing additional amino acids, such as, for example, an IgG Fc fusion region peptide. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

- 32 -

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. See Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).

In specific embodiments, the polypeptides of the invention comprise, or alternatively, consist of, fragments or variants of the amino acid sequence of a Therapeutic protein described herein and/or human serum albumin, and/or albumin fusion protein of the invention, wherein the fragments or variants have 1-5, 5-10, 5-25, 5-50, 10-50 or 50-150, amino acid residue additions, substitutions, and/or deletions when compared to the reference amino acid sequence. In certain embodiments, the amino acid substitutions are conservative. Nucleic acids encoding these polypeptides are also encompassed by the invention.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristylation, oxidation, pegylation, proteolytic

- 33 -

processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Furthermore, chemical entities may be covalently attached to the albumin fusion proteins to enhance or modulate a specific functional or biological activity such as by methods disclosed in Current Opinions in Biotechnology, 10:324 (1999).

Furthermore, targeting entities may be covalently attached to the albumin fusion proteins of the invention to target a specific functional or biological activity to certain cell or stage specific types, tissue types or anatomical structures. By directing albumin fusion proteins of the invention the action of the agent may be localized. Further, such targeting may enable the dosage of the albumin fusion proteins of the invention required to be reduced since, by accumulating the albumin fusion proteins of the invention at the required site, a higher localized concentration may be achieved. Albumin fusion proteins of the invention can be conjugated with a targeting portion by use of cross-linking agents as well as by recombinant DNA techniques whereby the nucleotide sequence encoding the albumin fusion proteins of the invention, or a functional portion of it, is cloned adjacent to the nucleotide sequence of the ligand when the ligand is a protein, and the conjugate expressed as a fusion protein.

Additional post-translational modifications encompassed by the invention include, for example, e.g., N-linked or O-linked carbohydrate chains, processing of N-terminal or C-terminal ends, attachment of chemical moieties to the amino acid backbone, chemical modifications of N-linked or O-linked carbohydrate chains, and addition or deletion of an N-terminal methionine residue as a result of procaryotic host cell expression. The albumin fusion proteins may also be modified with a detectable label, such as an enzymatic, fluorescent, isotopic or affinity label to allow for detection and isolation of the protein. Examples of such modifications are given, e.g., in U.S. Provisional Application Ser. No. 60/355,547 and in WO 01/79480 (pp. 105-106), which are incorporated by reference herein, and are well known in the art.

### **Functional activity**

“A polypeptide having functional activity” refers to a polypeptide capable of displaying one or more known functional activities associated with the full-length, pro-protein, and/or mature form of a Therapeutic protein. Such functional activities include, but are not limited to, biological activity, enzyme inhibition, antigenicity [ability to bind to an anti-polypeptide antibody or compete with a polypeptide for binding], immunogenicity (ability to generate an antibody which binds to a specific polypeptide of the invention), ability

- 34 -

to form multimers with polypeptides of the invention, and ability to bind to a receptor or ligand for a polypeptide.

“A polypeptide having biological activity” refers to a polypeptide exhibiting activity similar to, but not necessarily identical to, an activity of a Therapeutic protein of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention.

In other embodiments, an albumin fusion protein of the invention has at least one biological and/or therapeutic activity associated with the Therapeutic protein (or fragment or variant thereof) when it is not fused to albumin.

The albumin fusion proteins of the invention can be assayed for functional activity (e.g., biological activity) using or routinely modifying assays known in the art, as well as assays described herein. Specifically, albumin fusion proteins may be assayed for functional activity (e.g., biological activity or therapeutic activity) using the assay referenced in the “Relevant Publications” column of Table 4. Additionally, one of skill in the art may routinely assay fragments of a Therapeutic protein corresponding to a Therapeutic protein portion of an albumin fusion protein of the invention, for activity using assays referenced in its corresponding row of Table 4. Further, one of skill in the art may routinely assay fragments of an albumin protein corresponding to an albumin protein portion of an albumin fusion protein of the invention, for activity using assays known in the art and/or as described in the Examples section below.

In addition, assays described herein (see Examples and Table 4) and otherwise known in the art may routinely be applied to measure the ability of albumin fusion proteins of the present invention and fragments, variants and derivatives thereof to elicit biological activity and/or Therapeutic activity (either *in vitro* or *in vivo*) related to either the Therapeutic protein portion and/or albumin portion of the albumin fusion protein of the present invention. Other methods will be known to the skilled artisan and are within the scope of the invention.

#### **Expression of Fusion Proteins**

The albumin fusion proteins of the invention may be produced as recombinant molecules by secretion from yeast, a microorganism such as a bacterium, or a human or animal cell line. Optionally, the polypeptide is secreted from the host cells.

- 35 -

For expression of the albumin fusion proteins exemplified herein, yeast strains disrupted of the *HSP150* gene as exemplified in WO 95/33833, or yeast strains disrupted of the *PMT1* gene as exemplified in WO 00/44772 [rHA process] (serving to reduce/eliminate O-linked glycosylation of the albumin fusions), or yeast strains disrupted of the *YAP3* gene as exemplified in WO 95/23857 were successfully used, in combination with the yeast *PRB1* promoter, the HSA/MF $\alpha$ -1 fusion leader sequence exemplified in WO 90/01063, the yeast *ADHI* terminator, the *LEU2* selection marker and the disintegration vector pSAC35 exemplified in U.S. Patent No. 5,637,504.

Other yeast strains, promoters, leader sequences, terminators, markers and vectors which are expected to be useful in the invention are described in U.S. Provisional Application Serial No. 60/355,547 and in WO 01/74980 (pp. 94-99), which are incorporated herein by reference, and are well known in the art.

The present invention also includes a cell, optionally a yeast cell transformed to express an albumin fusion protein of the invention. In addition to the transformed host cells themselves, the present invention also contemplates a culture of those cells, optionally a monoclonal (clonally homogeneous) culture, or a culture derived from a monoclonal culture, in a nutrient medium. If the polypeptide is secreted, the medium will contain the polypeptide, with the cells, or without the cells if they have been filtered or centrifuged away. Many expression systems are known and may be used, including bacteria (for example *E. coli* and *Bacillus subtilis*), yeasts (for example *Saccharomyces cerevisiae*, *Kluyveromyces lactis* and *Pichia pastoris*), filamentous fungi (for example *Aspergillus*), plant cells, animal cells and insect cells.

The desired protein is produced in conventional ways, for example from a coding sequence inserted in the host chromosome or on a free plasmid. The yeasts are transformed with a coding sequence for the desired protein in any of the usual ways, for example electroporation. Methods for transformation of yeast by electroporation are disclosed in Becker & Guarente (1990) *Methods Enzymol.* 194, 182.

Successfully transformed cells, *i.e.*, cells that contain a DNA construct of the present invention, can be identified by well known techniques. For example, cells resulting from the introduction of an expression construct can be grown to produce the desired polypeptide. Cells can be harvested and lysed and their DNA content examined for the presence of the DNA using a method such as that described by Southern (1975) *J. Mol. Biol.* 98, 503 or

- 36 -

Berent *et al.* (1985) *Biotech.* 3, 208. Alternatively, the presence of the protein in the supernatant can be detected using antibodies.

Useful yeast plasmid vectors include pRS403-406 and pRS413-416 and are generally available from Stratagene Cloning Systems, La Jolla, CA 92037, USA. Plasmids pRS403, pRS404, pRS405 and pRS406 are Yeast Integrating plasmids (YIps) and incorporate the yeast selectable markers *HIS3*, *TRP1*, *LEU2* and *URA3*. Plasmids pRS413-416 are Yeast Centromere plasmids (YCps).

Vectors for making albumin fusion proteins for expression in yeast include pPPC0005, pScCHSA, pScNHSA, and pC4:HSA which were deposited on April 11, 2001 at the American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110-2209 and which are described in Provisional Application Serial No. 60/355,547 and WO 01/79480, which are incorporated by reference herein.

Another vector which is expected to be useful for expressing an albumin fusion protein in yeast is the pSAC35 vector which is described in Sleep *et al.*, *BioTechnology* 8:42 (1990), which is hereby incorporated by reference in its entirety. The plasmid pSAC35 is of the disintegration class of vector described in US 5,637,504.

A variety of methods have been developed to operably link DNA to vectors via complementary cohesive termini. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

Synthetic linkers containing one or more restriction sites provide an alternative method of joining the DNA segment to vectors. The DNA segment, generated by endonuclease restriction digestion, is treated with bacteriophage T4 DNA polymerase or *E. coli* DNA polymerase I, enzymes that remove protruding,  $\gamma$ -single-stranded termini with their 3' 5'-exonucleolytic activities, and fill in recessed 3'-ends with their polymerizing activities. The combination of these activities therefore generates blunt-ended DNA segments. The blunt-ended segments are then incubated with a large molar excess of linker molecules in the presence of an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules, such as bacteriophage T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying polymeric linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction enzyme and ligated to an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the DNA segment.

- 37 -

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of commercial sources.

A desirable way to modify the DNA in accordance with the invention, if, for example, HA variants are to be prepared, is to use the polymerase chain reaction as disclosed by Saiki *et al.* (1988) *Science* 239, 487-491. In this method the DNA to be enzymatically amplified is flanked by two specific oligonucleotide primers which themselves become incorporated into the amplified DNA. The specific primers may contain restriction endonuclease recognition sites which can be used for cloning into expression vectors using methods known in the art.

Exemplary genera of yeast contemplated to be useful in the practice of the present invention as hosts for expressing the albumin fusion proteins are *Pichia* (formerly classified as *Hansenula*), *Saccharomyces*, *Kluyveromyces*, *Aspergillus*, *Candida*, *Torulopsis*, *Torulaspora*, *Schizosaccharomyces*, *Citeromyces*, *Pachysolen*, *Zygosaccharomyces*, *Debaromyces*, *Trichoderma*, *Cephalosporium*, *Humicola*, *Mucor*, *Neurospora*, *Yarrowia*, *Metschnikowia*, *Rhodospiridium*, *Leucosporidium*, *Botryosphaeria*, *Sporidiobolus*, *Endomycopsis*, and the like. Genera include those selected from the group consisting of *Saccharomyces*, *Schizosaccharomyces*, *Kluyveromyces*, *Pichia* and *Torulaspora*. Examples of *Saccharomyces* spp. are *S. cerevisiae*, *S. italicus* and *S. rouxii*. Examples of other species, and methods of transforming them, are described in U.S. Provisional Application Serial No. 60/355,547 and WO 01/79480 (pp. 97-98), which are incorporated herein by reference.

Methods for the transformation of *S. cerevisiae* are taught generally in EP 251 744, EP 258 067 and WO 90/01063, all of which are incorporated herein by reference.

Suitable promoters for *S. cerevisiae* include those associated with the *PGKI* gene, *GAL1* or *GAL10* genes, *CYCI*, *PHO5*, *TRPI*, *ADHI*, *ADH2*, the genes for glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, triose phosphate isomerase, phosphoglucose isomerase, glucokinase, alpha-mating factor pheromone, [a mating factor pheromone], the *PRBI* promoter, the *GUT2* promoter, the *GPGI* promoter, and hybrid promoters involving hybrids of parts of 5' regulatory regions with parts of 5' regulatory regions of other promoters or with upstream activation sites (e.g. the promoter of EP-A-258 067).

Convenient regulatable promoters for use in *Schizosaccharomyces pombe* are the thiamine-repressible promoter from the *nmt* gene as described by Maundrell (1990) *J. Biol. Chem.* 265, 10857-10864 and the glucose repressible *jbpl* gene promoter as described by Hoffman & Winston (1990) *Genetics* 124, 807-816.

- 38 -

Methods of transforming *Pichia* for expression of foreign genes are taught in, for example, Cregg *et al.* (1993), and various Phillips patents (*e.g.* US 4 857 467, incorporated herein by reference), and *Pichia* expression kits are commercially available from Invitrogen BV, Leek, Netherlands, and Invitrogen Corp., San Diego, California. Suitable promoters include AOX1 and AOX2. Gleeson *et al.* (1986) J. Gen. Microbiol. 132, 3459-3465 include information on *Hansenula* vectors and transformation, suitable promoters being MOX1 and FMD1; whilst EP 361 991, Fleer *et al.* (1991) and other- publications from Rhone-Poulenc Rorer teach how to express foreign proteins in *Kluyveromyces* spp.

The transcription termination signal may be the 3' flanking sequence of a eukaryotic gene which contains proper signals for transcription termination and polyadenylation. Suitable 3' flanking sequences may, for example, be those of the gene naturally linked to the expression control sequence used, *i.e.* may correspond to the promoter. Alternatively, they may be different in which case the termination signal of the *S. cerevisiae* *ADHI* gene is optionally used.

The desired albumin fusion protein may be initially expressed with a secretion leader sequence, which may be any leader effective in the yeast chosen. Leaders useful in *S. cerevisiae* include that from the mating factor  $\alpha$  polypeptide (MF  $\alpha$ -1) and the hybrid leaders of EP-A-387 319. Such leaders (or signals) are cleaved by the yeast before the mature albumin is released into the surrounding medium. Further such leaders include those of *S. cerevisiae* invertase (*SUC2*) disclosed in JP 62-096086 (granted as 911036516), acid phosphatase (*PH05*), the pre-sequence of MF $\alpha$ -1,  $\alpha$  glucanase (*BGL2*) and killer toxin; *S. diastaticus* glucoamylase II; *S. carlsbergensis*  $\alpha$ -galactosidase (MEL1); *K. lactis* killer toxin; and *Candida* glucoamylase.

#### Additional Methods of Recombinant and Synthetic Production of Albumin Fusion Proteins

The present invention includes polynucleotides encoding albumin fusion proteins of this invention, as well as vectors, host cells and organisms containing these polynucleotides. The present invention also includes methods of producing albumin fusion proteins of the invention by synthetic and recombinant techniques. The polynucleotides, vectors, host cells, and organisms may be isolated and purified by methods known in the art.

A vector useful in the invention may be, for example, a phage, plasmid, cosmid, mini-chromosome, viral or retroviral vector.

The vectors which can be utilized to clone and/or express polynucleotides of the invention are vectors which are capable of replicating and/or expressing the polynucleotides



- 39 -

in the host cell in which the polynucleotides are desired to be replicated and/or expressed. In general, the polynucleotides and/or vectors can be utilized in any cell, either eukaryotic or prokaryotic, including mammalian cells (e.g., human (e.g., HeLa), monkey (e.g., Cos), rabbit (e.g., rabbit reticulocytes), rat, hamster (e.g., CHO, NSO and baby hamster kidney cells) or mouse cells (e.g., L cells), plant cells, yeast cells, insect cells or bacterial cells (e.g., *E. coli*). See, e.g., F. Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience (1992) and Sambrook et al. (1989) for examples of appropriate vectors for various types of host cells. Note, however, that when a retroviral vector that is replication defective is used, viral propagation generally will occur only in complementing host cells.

The host cells containing these polynucleotides can be used to express large amounts of the protein useful in, for example, pharmaceuticals, diagnostic reagents, vaccines and therapeutics. The protein may be isolated and purified by methods known in the art or described herein.

The polynucleotides encoding albumin fusion proteins of the invention may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector may be introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter compatible with the host cell in which the polynucleotide is to be expressed. The promoter may be a strong promoter and/or an inducible promoter. Examples of promoters include the phage lambda PL promoter, the *E. coli lac*, *trp*, *phoA* and *tac* promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs may include a translation initiating codon at the beginning and a termination codon (TAA, TGA or TAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors may include at least one selectable marker. Such markers include dihydrofolate reductase, G418, glutamine synthase, or neomycin resistance for eukaryotic cell culture, and tetracycline, kanamycin or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella*

- 40 -

*typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* or *Pichia pastoris* (ATCC Accession No. 201178)); insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, NSO, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

In one embodiment, polynucleotides encoding an albumin fusion protein of the invention may be fused to signal sequences which will direct the localization of a protein of the invention to particular compartments of a prokaryotic or eukaryotic cell and/or direct the secretion of a protein of the invention from a prokaryotic or eukaryotic cell. For example, in *E. coli*, one may wish to direct the expression of the protein to the periplasmic space. Examples of signal sequences or proteins (or fragments thereof) to which the albumin fusion proteins of the invention may be fused in order to direct the expression of the polypeptide to the periplasmic space of bacteria include, but are not limited to, the *pelB* signal sequence, the maltose binding protein (MBP) signal sequence, MBP, the *ompA* signal sequence, the signal sequence of the periplasmic *E. coli* heat-labile enterotoxin B-subunit, and the signal sequence of alkaline phosphatase. Several vectors are commercially available for the construction of fusion proteins which will direct the localization of a protein, such as the pMAL series of vectors (particularly the pMAL-p series) available from New England Biolabs. In a specific embodiment, polynucleotides albumin fusion proteins of the invention may be fused to the *pelB* pectate lyase signal sequence to increase the efficiency of expression and purification of such polypeptides in Gram-negative bacteria. See, U.S. Patent Nos. 5,576,195 and 5,846,818, the contents of which are herein incorporated by reference in their entireties.

Examples of signal peptides that may be fused to an albumin fusion protein of the invention in order to direct its secretion in mammalian cells include, but are not limited to, the MPIF-1 signal sequence (e.g., amino acids 1-21 of GenBank Accession number AAB51134), the stanniocalcin signal sequence (MLQNSAVLLLLVISASA, SEQ ID NO:\_\_, and a consensus signal sequence (MPTWAWWLFLVLLLALWAPARG, SEQ ID NO:\_\_. A suitable signal sequence that may be used in conjunction with baculoviral expression systems is the gp67 signal sequence (e.g., amino acids 1-19 of GenBank Accession Number AAA72759).

Vectors which use glutamine synthase (GS) or DHFR as the selectable markers can be amplified in the presence of the drugs methionine sulfoximine or methotrexate, respectively. An advantage of glutamine synthase based vectors is the availability of cell lines (e.g., the murine myeloma cell line, NSO) which are glutamine synthase negative. Glutamine synthase

- 41 -

expression systems can also function in glutamine synthase expressing cells (e.g., Chinese Hamster Ovary (CHO) cells) by providing additional inhibitor to prevent the functioning of the endogenous gene. A glutamine synthase expression system and components thereof are detailed in PCT publications: WO87/04462; WO86/05807; WO89/01036; WO89/10404; and WO91/06657, which are hereby incorporated in their entireties by reference herein. Additionally, glutamine synthase expression vectors can be obtained from Lonza Biologics, Inc. (Portsmouth, NH). Expression and production of monoclonal antibodies using a GS expression system in murine myeloma cells is described in Bebbington *et al.*, *Bio/technology* 10:169(1992) and in Biblia and Robinson *Biotechnol. Prog.* 11:1 (1995) which are herein incorporated by reference.

The present invention also relates to host cells containing vector constructs, such as those described herein, and additionally encompasses host cells containing nucleotide sequences of the invention that are operably associated with one or more heterologous control regions (e.g., promoter and/or enhancer) using techniques known of in the art. The host cell can be a higher eukaryotic cell, such as a mammalian cell (e.g., a human derived cell), or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. A host strain may be chosen which modulates the expression of the inserted gene sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus expression of the genetically engineered polypeptide may be controlled. Furthermore, different host cells have characteristics and specific mechanisms for the translational and post-translational processing and modification (e.g., phosphorylation, cleavage) of proteins. Appropriate cell lines can be chosen to ensure the desired modifications and processing of the foreign protein expressed.

Introduction of the nucleic acids and nucleic acid constructs of the invention into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis *et al.*, *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

In addition to encompassing host cells containing the vector constructs discussed herein, the invention also encompasses primary, secondary, and immortalized host cells of vertebrate origin, particularly mammalian origin, that have been engineered to delete or

- 42 -

replace endogenous genetic material (e.g., the coding sequence corresponding to a Therapeutic protein may be replaced with an albumin fusion protein corresponding to the Therapeutic protein), and/or to include genetic material (e.g., heterologous polynucleotide sequences such as for example, an albumin fusion protein of the invention corresponding to the Therapeutic protein may be included). The genetic material operably associated with the endogenous polynucleotide may activate, alter, and/or amplify endogenous polynucleotides.

In addition, techniques known in the art may be used to operably associate heterologous polynucleotides (e.g., polynucleotides encoding an albumin protein, or a fragment or variant thereof) and/or heterologous control regions (e.g., promoter and/or enhancer) with endogenous polynucleotide sequences encoding a Therapeutic protein via homologous recombination (see, e.g., US Patent Number 5,641,670, issued June 24, 1997; International Publication Number WO 96/29411; International Publication Number WO 94/12650; Koller *et al.*, *Proc. Natl. Acad. Sci. USA* 86:8932-8935 (1989); and Zijlstra *et al.*, *Nature* 342:435-438 (1989), the disclosures of each of which are incorporated by reference in their entireties).

Advantageously, albumin fusion proteins of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, hydrophobic charge interaction chromatography and lectin chromatography. In some embodiments, high performance liquid chromatography ("HPLC") may be employed for purification. In some cases, therapeutic proteins have low solubility or are soluble only in low or high pH or only in high or low salt. Fusion of therapeutic proteins to HSA is likely to improve the solubility characteristics of the therapeutic protein.

In some embodiments albumin fusion proteins of the invention are purified using one or more Chromatography methods listed above. In other embodiments, albumin fusion proteins of the invention are purified using one or more of the following Chromatography columns, Q sepharose FF column, SP Sepharose FF column, Q Sepharose High Performance Column, Blue Sepharose FF column, Blue Column, Phenyl Sepharose FF column, DEAE Sepharose FF, or Methyl Column.

Additionally, albumin fusion proteins of the invention may be purified using the process described in International Publication No. WO 00/44772 which is herein incorporated

- 43 -

by reference in its entirety. One of skill in the art could easily modify the process described therein for use in the purification of albumin fusion proteins of the invention.

Albumin fusion proteins of the present invention may be recovered from products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, albumin fusion proteins of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Albumin fusion proteins of the invention and antibodies that bind a Therapeutic protein or fragments or variants thereof can be fused to marker sequences, such as a peptide to facilitate purification. In one embodiment, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Other peptide tags useful for purification include, but are not limited to, the "HA" tag, which corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson et al., Cell 37:767 (1984)) and the "FLAG" tag.

Further, an albumin fusion protein of the invention may be conjugated to a therapeutic moiety such as a cytotoxin, e.g., a cytostatic or cytocidal agent, a therapeutic agent or a radioactive metal ion, e.g., alpha-emitters such as, for example, <sup>213</sup>Bi. Examples of such agents are given in U.S. Provisional Application Serial No. 60/355,547 and in WO 01/79480 (p. 107), which are incorporated herein by reference.

Albumin fusion proteins may also be attached to solid supports, which are particularly useful for immunoassays or purification of polypeptides that are bound by, that bind to, or associate with albumin fusion proteins of the invention. Such solid supports include, but are not limited to, glass, cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene.

- 44 -

Also provided by the invention are chemically modified derivatives of the albumin fusion proteins of the invention which may provide additional advantages such as increased solubility, stability and circulating time of the polypeptide, or decreased immunogenicity (see U.S. Patent No. 4,179,337). Examples involving the use of polyethylene glycol are given in WO 01/79480 (pp. 109-111), which are incorporated by reference herein.

The presence and quantity of albumin fusion proteins of the invention may be determined using ELISA, a well known immunoassay known in the art.

### **Uses of the Polypeptides**

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

The albumin fusion proteins of the present invention are useful for treatment, prevention and/or prognosis of various disorders in mammals, preferably humans. Such disorders include, but are not limited to, those described herein under the heading "Biological Activity" in Table 4. For example, the albumin fusion proteins of the present invention may be used as inhibitors of serine proteases, plasmin, human neutrophil elastase and/or kallikrein.

Albumin fusion proteins can also be used to assay levels of polypeptides in a biological sample. For example, radiolabeled albumin fusion proteins of the invention could be used for imaging of polypeptides in a body. Examples of assays are given, e.g., in U.S. Provisional Application Serial No. 60/355,547 and WO 0179480 (pp. 112-122), which are incorporated herein by reference, and are well known in the art. Labels or markers for *in vivo* imaging of protein include, but are not limited to, those detectable by X-radiography, nuclear magnetic resonance (NMR), electron spin relaxation (ESR), positron emission tomography (PET), or computer tomography (CT). For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the albumin fusion protein by labeling of nutrients given to a cell line expressing the albumin fusion protein of the invention.

An albumin fusion protein which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example,  $^{131}\text{I}$ ,  $^{112}\text{In}$ ,  $^{99\text{m}}\text{Tc}$ , ( $^{131}\text{I}$ ,  $^{125}\text{I}$ ,  $^{123}\text{I}$ ,  $^{121}\text{I}$ ), carbon ( $^{14}\text{C}$ ), sulfur ( $^{35}\text{S}$ ), tritium ( $^3\text{H}$ ), indium ( $^{115\text{m}}\text{In}$ ,  $^{113\text{m}}\text{In}$ ,  $^{112}\text{In}$ ,  $^{111}\text{In}$ ), and technetium ( $^{99}\text{Tc}$ ,  $^{99\text{m}}\text{Tc}$ ), thallium ( $^{201}\text{Tl}$ ), gallium ( $^{68}\text{Ga}$ ,  $^{67}\text{Ga}$ ), palladium ( $^{103}\text{Pd}$ ), molybdenum ( $^{99}\text{Mo}$ ), xenon ( $^{133}\text{Xe}$ ), fluorine ( $^{18}\text{F}$ ,  $^{153}\text{Sm}$ ,  $^{177}\text{Lu}$ ,  $^{159}\text{Gd}$ ,  $^{149}\text{Pm}$ ,  $^{140}\text{La}$ ,  $^{175}\text{Yb}$ ,  $^{166}\text{Ho}$ ,  $^{90}\text{Y}$ ,  $^{47}\text{Sc}$ ,  $^{186}\text{Re}$ ,

- 45 -

$^{188}\text{Re}$ ,  $^{142}\text{Pr}$ ,  $^{105}\text{Rh}$ ,  $^{97}\text{Ru}$ ), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously or intraperitoneally) into the mammal to be examined for immune system disorder. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of  $^{99\text{m}}\text{Tc}$ . The labeled albumin fusion protein will then preferentially accumulate at locations in the body (e.g., organs, cells, extracellular spaces or matrices) where one or more receptors, ligands or substrates (corresponding to that of the Therapeutic protein used to make the albumin fusion protein of the invention) are located. Alternatively, in the case where the albumin fusion protein comprises at least a fragment or variant of a Therapeutic antibody, the labeled albumin fusion protein will then preferentially accumulate at the locations in the body (e.g., organs, cells, extracellular spaces or matrices) where the polypeptides/epitopes corresponding to those bound by the Therapeutic antibody (used to make the albumin fusion protein of the invention) are located. *In vivo* tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments" (Chapter 13 in *Tumor Imaging: The Radiochemical Detection of Cancer*, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982)). The protocols described therein could easily be modified by one of skill in the art for use with the albumin fusion proteins of the invention.

Albumin fusion proteins of the invention can also be used to raise antibodies, which in turn may be used to measure protein expression of the Therapeutic protein, albumin protein, and/or the albumin fusion protein of the invention from a recombinant cell, as a way of assessing transformation of the host cell, or in a biological sample. Moreover, the albumin fusion proteins of the present invention can be used to test the biological activities described herein.

### **Transgenic Organisms**

Transgenic organisms that express the albumin fusion proteins of the invention are also included in the invention. Transgenic organisms are genetically modified organisms into which recombinant, exogenous or cloned genetic material has been transferred. Such genetic material is often referred to as a transgene. The nucleic acid sequence of the transgene may include one or more transcriptional regulatory sequences and other nucleic acid sequences such as introns, that may be necessary for optimal expression and secretion of the encoded

- 46 -

protein. The transgene may be designed to direct the expression of the encoded protein in a manner that facilitates its recovery from the organism or from a product produced by the organism, *e.g.* from the milk, blood, urine, eggs, hair or seeds of the organism. The transgene may consist of nucleic acid sequences derived from the genome of the same species or of a different species than the species of the target animal. The transgene may be integrated either at a locus of a genome where that particular nucleic acid sequence is not otherwise normally found or at the normal locus for the transgene.

The term "germ cell line transgenic organism" refers to a transgenic organism in which the genetic alteration or genetic information was introduced into a germ line cell, thereby conferring the ability of the transgenic organism to transfer the genetic information to offspring. If such offspring in fact possess some or all of that alteration or genetic information, then they too are transgenic organisms. The alteration or genetic information may be foreign to the species of organism to which the recipient belongs, foreign only to the particular individual recipient, or may be genetic information already possessed by the recipient. In the last case, the altered or introduced gene may be expressed differently than the native gene.

A transgenic organism may be a transgenic human, animal or plant. Transgenics can be produced by a variety of different methods including transfection, electroporation, microinjection, gene targeting in embryonic stem cells and recombinant viral and retroviral infection (*see, e.g.*, U.S. Patent No. 4,736,866; U.S. Patent No. 5,602,307; Mullins *et al.* (1993) *Hypertension* 22(4):630-633; Brenin *et al.* (1997) *Surg. Oncol.* 6(2)99-110; Tuan (ed.), *Recombinant Gene Expression Protocols*, Methods in Molecular Biology No. 62, Humana Press (1997)). The method of introduction of nucleic acid fragments into recombination competent mammalian cells can be by any method which favors co-transformation of multiple nucleic acid molecules. Detailed procedures for producing transgenic animals are readily available to one skilled in the art, including the disclosures in U.S. Patent No. 5,489,743 and U.S. Patent No. 5,602,307. Additional information is given in U.S. Provisional Application Serial No. 60/355,547 and WO 01/79480 (pp. 151-162), which are incorporated by reference herein.

### **Gene Therapy**

Constructs encoding albumin fusion proteins of the invention can be used as a part of a gene therapy protocol to deliver therapeutically effective doses of the albumin fusion protein. One approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector



containing nucleic acid, encoding an albumin fusion protein of the invention. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral vector, *e.g.*, by a cDNA contained in the viral vector, are expressed efficiently in cells which have taken up viral vector nucleic acid. The extended plasma half-life of the described albumin fusion proteins may even compensate for a potentially low expression level.

Retrovirus vectors and adeno-associated virus vectors can be used as a recombinant gene delivery system for the transfer of exogenous nucleic acid molecules encoding albumin fusion proteins *in vivo*. These vectors provide efficient delivery of nucleic acids into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host. Examples of such vectors, methods of using them, and their advantages, as well as non-viral delivery methods are described in detail in U.S. Provisional Application Serial No. 60/355,547 and WO 01/79480 (pp. 151-153), which are incorporated by reference herein.

Gene delivery systems for a gene encoding an albumin fusion protein of the invention can be introduced into a patient by any of a number of methods. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, *e.g.* by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by Stereotactic injection (*e.g.* Chen *et al.* (1994) *PNAS* 91: 3054-3057). The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Where the albumin fusion protein can be produced intact from recombinant cells, *e.g.* retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the albumin fusion protein. Additional gene therapy methods are described in U.S. Provisional Application Serial No. 60/355,547 and in WO 01/79480 (pp. 153-162), which are incorporated herein by reference.

### **Pharmaceutical or Therapeutic Compositions**

The albumin fusion proteins of the invention or formulations thereof may be administered by any conventional method including parenteral (*e.g.* subcutaneous or

- 48 -

intramuscular) injection or intravenous infusion. The treatment may consist of a single dose or a plurality of doses over a period of time. Furthermore, the dose, or plurality of doses, is administered less frequently than for the Therapeutic Protein which is not fused to albumin.

While it is possible for an albumin fusion protein of the invention to be administered alone, it is desirable to present it as a pharmaceutical formulation, together with one or more acceptable carriers. The carrier(s) must be "acceptable" in the sense of being compatible with the albumin fusion protein and not deleterious to the recipients thereof. Typically, the carriers will be water or saline which will be sterile and pyrogen free. Albumin fusion proteins of the invention are particularly well suited to formulation in aqueous carriers such as sterile pyrogen free water, saline or other isotonic solutions because of their extended shelf-life in solution. For instance, pharmaceutical compositions of the invention may be formulated well in advance in aqueous form, for instance, weeks or months or longer time periods before being dispensed.

Formulations containing the albumin fusion protein may be prepared taking into account the extended shelf-life of the albumin fusion protein in aqueous formulations. As discussed above, the shelf-life of many of these Therapeutic proteins are markedly increased or prolonged after fusion to HA.

In instances where aerosol administration is appropriate, the albumin fusion proteins of the invention can be formulated as aerosols using standard procedures. The term "aerosol" includes any gas-borne suspended phase of an albumin fusion protein of the instant invention which is capable of being inhaled into the bronchioles or nasal passages. Specifically, aerosol includes a gas-borne suspension of droplets of an albumin fusion protein of the instant invention, as may be produced in a metered dose inhaler or nebulizer, or in a mist sprayer. Aerosol also includes a dry powder composition of a compound of the instant invention suspended in air or other carrier gas, which may be delivered by insufflation from an inhaler device, for example.

The formulations may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. Such methods include the step of bringing into association the albumin fusion protein with the carrier that constitutes one or more accessory ingredients. In general the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes

- 49 -

which render the formulation appropriate for the intended recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example sealed ampules, vials or syringes, and may be stored in a freeze-dried (lyophilised) condition requiring only the addition of the sterile liquid carrier, for example water for injections, immediately prior to use. Extemporaneous injection solutions and suspensions may be prepared from sterile powders. Dosage formulations may contain the Therapeutic protein portion at a lower molar concentration or lower dosage compared to the non-fused standard formulation for the Therapeutic protein given the extended serum half-life exhibited by many of the albumin fusion proteins of the invention.

As an example, when an albumin fusion protein of the invention comprises one or more of the Therapeutic protein regions, the dosage form can be calculated on the basis of the potency of the albumin fusion protein relative to the potency of the Therapeutic protein, while taking into account the prolonged serum half-life and shelf-life of the albumin fusion proteins compared to that of the native Therapeutic protein. For example, in an albumin fusion protein consisting of a full length HA fused to a full length Therapeutic protein, an equivalent dose in terms of units would represent a greater weight of agent but the dosage frequency can be reduced.

Formulations or compositions of the invention may be packaged together with, or included in a kit with, instructions or a package insert referring to the extended shelf-life of the albumin fusion protein component. For instance, such instructions or package inserts may address recommended storage conditions, such as time, temperature and light, taking into account the extended or prolonged shelf-life of the albumin fusion proteins of the invention. Such instructions or package inserts may also address the particular advantages of the albumin fusion proteins of the inventions, such as the ease of storage for formulations that may require use in the field, outside of controlled hospital, clinic or office conditions. As described above, formulations of the invention may be in aqueous form and may be stored under less than ideal circumstances without significant loss of therapeutic activity.

The invention also provides methods of treatment and/or prevention of diseases or disorders (such as, for example, any one or more of the diseases or disorders disclosed herein) by administration to a subject of an effective amount of an albumin fusion protein of the invention or a polynucleotide encoding an albumin fusion protein of the invention ("albumin fusion polynucleotide") in a pharmaceutically acceptable carrier.

- 50 -

Effective dosages of the albumin fusion protein and/or polynucleotide of the invention to be administered may be determined through procedures well known to those in the art which address such parameters as biological half-life, bioavailability, and toxicity, including using data from routine *in vitro* and *in vivo* studies such as those described in the references in Table 4, using methods well known to those skilled in the art.

The albumin fusion protein and/or polynucleotide will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the albumin fusion protein and/or polynucleotide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

For example, determining an effective amount of substance to be delivered can depend upon a number of factors including, for example, the chemical structure and biological activity of the substance, the age and weight of the animal, the precise condition requiring treatment and its severity, and the route of administration. The frequency of treatments depends upon a number of factors, such as the amount of polynucleotide constructs administered per dose, as well as the health and history of the subject. The precise amount, number of doses, and timing of doses will be determined by the attending physician or veterinarian.

Albumin fusion proteins and polynucleotides of the present invention can be administered to any animal, preferably to mammals and birds. Preferred mammals include humans, dogs, cats, mice, rats, rabbits sheep, cattle, horses and pigs, with humans being particularly preferred.

As a general proposition, the albumin fusion protein of the invention will be dosed lower or administered less frequently than the unfused Therapeutic peptide. A therapeutically effective dose may refer to that amount of the compound sufficient to result in amelioration of symptoms, disease stabilization, a prolongation of survival in a patient, or improvement in the quality of life.

Albumin fusion proteins and/or polynucleotides can be administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any. The term "parenteral" as used

- 51 -

herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

Albumin fusion proteins and/or polynucleotides of the invention are also suitably administered by sustained-release systems, such as those described in U.S. Provisional Application Serial No. 60/355,547 and WO 01/79480 (pp. 129-130), which are incorporated by reference herein.

For parenteral administration, in one embodiment, the albumin fusion protein and/or polynucleotide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation optionally does not include oxidizing agents and other compounds that are known to be deleterious to the Therapeutic.

The albumin fusion proteins and/or polynucleotides of the invention may be administered alone or in combination with other therapeutic agents. Albumin fusion protein and/or polynucleotide agents that may be administered in combination with the albumin fusion proteins and/or polynucleotides of the invention, include but not limited to, chemotherapeutic agents, antibiotics, steroidal and non-steroidal anti-inflammatories, conventional immunotherapeutic agents, and/or therapeutic treatments as described in U.S. Provisional Application Serial No. 60/355,547 and WO 01/79480 (pp. 132-151) which are incorporated by reference herein. Combinations may be administered either concomitantly, e.g., as an admixture, separately but simultaneously or concurrently; or sequentially. This includes presentations in which the combined agents are administered together as a therapeutic mixture, and also procedures in which the combined agents are administered separately but simultaneously, e.g., as through separate intravenous lines into the same individual. Administration "in combination" further includes the separate administration of one of the compounds or agents given first, followed by the second.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions comprising albumin fusion proteins of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the

- 52 -

manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

With this general description of the invention, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, make and utilize the alterations detected in the present invention and practice the claimed methods. The following working examples therefore, specifically point out different embodiments of the present invention, and are not to be construed as limiting in any way the remainder of the disclosure.

## **EXAMPLES**

### **Example 1: Construction of N-terminal and C-terminal albumin-(GGG)<sub>4</sub>GG linker cloning vectors**

The recombinant albumin expression vectors pDB2243 and pDB2244 have been described previously in patent application WO 00/44772. The recombinant albumin expression vectors pAYE645 and pAYE646 have been described previously in UK patent application 0217033.0. Plasmid pDB2243 was modified to introduce a DNA sequence encoding the 14 amino acid polypeptide linker N-GGSGGSGGSGGSGG-C ((GGG)<sub>4</sub>GG, "N" and "C" denote the orientation of the polypeptide sequence) (SEQ ID NO:\_\_\_) at the C-terminal end of the albumin polypeptide in such a way to subsequently enable another polypeptide chain to be inserted C-terminal to the (GGG)<sub>4</sub>GG linker to produce a C-terminal albumin fusion in the general configuration, albumin-(GGG)<sub>4</sub>GG-polypeptide. Similarly, plasmid pAYE645 was modified to introduce a DNA sequence encoding the (GGG)<sub>4</sub>GG polypeptide linker at the N-terminal end of the albumin polypeptide in such a way to subsequently enable another polypeptide chain to be inserted N-terminal to the (GGG)<sub>4</sub>GG linker to produce an N-terminal albumin fusion in the general configuration of polypeptide-(GGG)<sub>4</sub>GG-albumin.

Plasmid pDB2243, described by Sleep, D., *et al.* (1991) *Bio/Technology* 9, 183-187 and in patent application WO 00/44772 which contained the yeast *PRB1* promoter and the yeast *ADH1* terminator providing appropriate transcription promoter and transcription terminator sequences. Plasmid pDB2243 was digested to completion with *Bam*HI, the recessed ends were blunt ended with T4 DNA polymerase and dNTPs, and finally religated to generate plasmid pDB2566.

A double stranded synthetic oligonucleotide linker *Bsu*36I/*Hind*III linker was synthesized by annealing the synthetic oligonucleotides JH033A and JH033B.

- 53 -

**JH033A**

5-TTAGGCTTAGGTGGTTCTGGTGGTCCGGTGGTTCTGGTGG  
ATCCGGTGGTTAATA-3'

(SEQ ID NO:\_\_\_)

**JH033B**

5'-AGCTTATTAACCACCGGATCCACCAGAACCACCGGAACCA  
CCAGAACCACCTAAGCC-3'

(SEQ ID NO:\_\_\_)

The annealed *Bsu36I/HindIII* linker was ligated into *HindIII/Bsu36I* cut pDB2566 to generate plasmid pDB2575X which comprised an albumin coding region with a (GGG)<sub>4</sub>GG peptide linker at its C-terminal end.

Plasmid pAYE645 that contained the yeast *PRB1* promoter and the yeast *ADH1* terminator providing appropriate transcription promoter and transcription terminator sequences is described in UK patent application 0217033.0. Plasmid pAYE645 was digested to completion with the restriction enzyme *AflIII* and partially digested with the restriction enzyme *HindIII* and the DNA fragment comprising the 3' end of the yeast *PRB1* promoter and the rHA coding sequence was isolated. Plasmid pDB2241 described in patent application WO 00/44772, was digested with *AflIII/HindIII* and the DNA fragment comprising the 5' end of the yeast *PRB1* promoter and the yeast *ADH1* terminator was isolated. The *AflIII/HindIII* DNA fragment from pAYE645 was then cloned into the *AflIII/HindIII* pDB2241 vector DNA fragment to create the plasmid pDB2302. Plasmid pDB2302 was digested to completion with *PacI/XhoI* and the 6.19kb fragment isolated, the recessed ends were blunt ended with T4 DNA polymerase and dNTPs, and religated to generate plasmid pDB2465. Plasmid pDB2465 was linearized with *ClaI*, the recessed ends were blunt ended with T4 DNA polymerase and dNTPs, and religated to generate plasmid pDB2533. Plasmid pDB2533 was linearized with *BlnI*, the recessed ends were blunt ended with T4 DNA polymerase and dNTPs, and religated to generate plasmid pDB2534. Plasmid pDB2534 was digested to completion with *BmgBI/BglII*, the 6.96kb DNA fragment isolated and ligated to one of two double stranded oligonucleotide linkers, VC053/VC054 and VC057/VC058 to create plasmid pDB2540, or VC055/VC056 and VC057/VC058 to create plasmid pDB2541.

- 54 -

**VC053**

5'-GATCTTTGGATAAGAGAGACGCTCACAAGTCCGAAGTCGCTCACCGGT-3'

(SEQ ID NO:\_\_\_)

**VC054**

5'-pCCTTGAACCGGTGAGCGACTTCGGACTTGTGAGCGTCTCTTATCCAAA-3'

(SEQ ID NO:\_\_\_)

**VC055**

5'-GATCTTTGGATAAGAGAGACGCTCACAAGTCCGAAGTCGCTCATCGAT-3'

(SEQ ID NO:\_\_\_)

**VC056**

5'-pCCTTGAATCGATGAGCGACTTCGGACTTGTGAGCGTCTCTTATCCAAA-3'

(SEQ ID NO:\_\_\_)

**VC057**5'-pTCAAGGACCTAGGTGAGGAAAACCTCAAGGCTTTGGTCTTGATCGCTTTCG  
CTCAATACTTGCAACAATGTCCATTTCGAAGATCAC-3'

(SEQ ID NO:\_\_\_)

**VC058**5'-GTGATCTTCGAATGGACATTGTTGCAAGTATTGAGCGAAAGCGATCAAGACC  
AAAGCCTTGAAGTTTTCTCACCTAGGT-3'

(SEQ ID NO:\_\_\_)

A double stranded synthetic oligonucleotide linker BglII/AgeI linker was synthesized by annealing the synthetic oligonucleotides JH035A and JH035B.

**JH035A**

5'-GATCTTTGGATAAGAGAGGTGGATCCGGTGGTTCCGGTGGTTCTGGTGGTTCCG



- 55 -

GTGGTGACGCTCACAAGTCCGAAGTCGCTCA-3''

(SEQ ID NO:\_\_\_)

**JH035B**

5'-

CCGGTGAGCGACTTCGGACTTGTGAGCGTCACCACCGGAACCACCAGAACCACC  
GGAACCACCGGATCCACCTCTCTTATCCAAA-3'

(SEQ ID NO:\_\_\_)

The annealed *Bgl*III/*Age*I linker was ligated into *Bgl*III/*Age*I cut pDB2540 to generate plasmid pDB2573X, which comprised an albumin coding region with a (GGG)<sub>4</sub>GG peptide linker at its N-terminal end.

**Example 2: Equilibrium Inhibition Constant for Unfused DPI-14**

The amino acid sequence of DPI-14 is

EAVREVCSEQAETGPCIAFFPRWYFDVTEGKCAPFFYGGCGGNRNNFDTEEYCMVCGSA

(SEQ ID NO:\_\_\_). A DNA sequence was derived from this polypeptide sequence by the process of back-translation. The DPI-14 was expressed in *Pichia* and extracted from the fermentation broth supernatant using ion-exchange chromatography, hydrophobic interaction chromatography, and ultrafiltration. The equilibrium inhibition constant ( $K_i$ ) for DPI-14 inhibition of human neutrophil elastase (HNE) was determined to be  $15 \pm 2$  pM, for [HNE] =  $57 \pm 7$  pM. The  $K_i$  measurement was performed using the methods set forth in Example 15.

**Example 3: A Construction of N-terminal and C-terminal albumin-DPI-14 fusions**

The DNA sequences were provided at the 5' or 3' end to encode bridging sequences between the DPI-14 coding region, the albumin coding region or the leader sequence as appropriate for N-terminal DPI-14-(GGG)<sub>4</sub>GG-albumin or C-terminal albumin-(GGG)<sub>4</sub>GG-DPI-14 fusions. An N-terminal *Bgl*III-*Bam*HI DPI-14 cDNA (Table 5) and a C-terminal *Bam*HI-*Hind*III DPI-14 cDNA (Table 6) were constructed from overlapping oligonucleotides.

**Example 4: Construction of N-terminal DPI-14-(GGG)<sub>4</sub>GG-albumin expression plasmids**

Plasmid pDB2573X was digested to completion with *Bgl*III and *Bam*HI, the 6.21kb DNA fragment was isolated and treated with calf intestinal phosphatase and then ligated with the 0.2kb *Bgl*III/*Bam*HI N terminal DPI-14 cDNA to create pDB2666. The DNA and amino

- 56 -

acid sequence of the N-terminal DPI-14-(GGG)<sub>4</sub>GG-albumin fusion are shown in Table 7 and Table 8, respectively. Appropriate yeast vector sequences were provided by a "disintegration" plasmid pSAC35 generally disclosed in EP-A-286 424 and described by Sleep, D., *et al.* (1991) *Bio/Technology* 9, 183-187. The *NotI* N-terminal DPI-14-(GGG)<sub>4</sub>GG-rHA expression cassette was isolated from pDB2666, purified and ligated into *NotI* digested pSAC35 which had been treated with calf intestinal phosphatase, creating two plasmids; the first (pDB2679) contained the *NotI* expression cassette in the same expression orientation as *LEU2*, while the second (pDB2680) contained the *NotI* expression cassette in the opposite orientation to *LEU2*. Both pDB2679 and pDB2680 are good producers of the desired fusion protein.

**Example 5: Construction of C-terminal albumin-(GGG)<sub>4</sub>GG-DPI-14 expression plasmid**

Plasmid pDB2575X was partially digested with *HindIII* and then digested to completion with *BamHI*. The desired 6.55kb DNA fragment was isolated and ligated with the 0.2kb *BamHI/HindIII* C terminal DPI-14 cDNA to create pDB2648. The DNA and amino acid sequence of the C-terminal albumin-(GGG)<sub>4</sub>GG-DPI-14 fusion are shown in Table 9 and Table 10, respectively. Appropriate yeast vector sequences were provided by a "disintegration" plasmid pSAC35 generally disclosed in EP-A-286 424 and described by Sleep, D., *et al.* (1991) *Bio/Technology* 9, 183-187. The *NotI* C-terminal albumin-(GGG)<sub>4</sub>GG-DPI-14 expression cassette was isolated from pDB2648, purified and ligated into *NotI* digested pSAC35 which had been treated with calf intestinal phosphatase, creating pDB2651 contained the *NotI* expression cassette in the same expression orientation as *LEU2*.

**Example 6: Construction of C-terminal albumin-(GGG)<sub>4</sub>GG-DX-1000 expression plasmid**

Plasmid pDB2575X was partially digested with *HindIII* and then digested to completion with *BamHI*. The desired 6.55kb DNA fragment was isolated and ligated with the 0.2kb *BamHI/HindIII* C-terminal DX-1000 cDNA as shown in Table 11 to create pDB2648X-1000. Appropriate yeast vector sequences were provided by a "disintegration" plasmid pSAC35 generally disclosed in EP-A-286 424 and described by Sleep, D., *et al.* (1991) *Bio/Technology* 9, 183-187. The *NotI* C-terminal albumin-(GGG)<sub>4</sub>GG-DX1000 expression cassette was isolated from pDB2648X-1000, purified and ligated into *NotI* digested pSAC35 which had been treated with calf intestinal phosphatase, creating pDB2651X-1000 contained the *NotI* expression cassette in the same expression orientation as *LEU2*.

**Example 7: Construction of N-terminal and C-terminal albumin-DX-890 fusions****Generation of the basic clone**

The amino acid sequence of DX-890 is  
 EACNLPIVRGPCIAFFPRWAFDAVKGKCVLFPYGGCQGNGNKFYSEKECREYCGVP  
 (SEQ ID NO: \_\_\_\_). A DNA sequence was derived from this polypeptide sequence by the process of back-translation. The DNA sequences were provided at the 5' or 3' end to encode bridging sequences between the DX-890 coding region, the albumin coding region or the leader sequence as appropriate for N-terminal DX-890-(GGG)<sub>4</sub>GG-albumin or C-terminal albumin-(GGG)<sub>4</sub>GG-DX-890 fusions. An N-terminal *Bgl*III-*Bam*HI DX-890 cDNA (Table 12) and a C-terminal *Bam*HI-*Hind*III DX-890 cDNA (Table 13) were constructed from overlapping oligonucleotides.

**Example 8: Construction of N-terminal DX-890-(GGG)<sub>4</sub>GG-albumin expression plasmids**

Plasmid pDB2573X was digested to completion with *Bgl*III and *Bam*HI, the 6.21kb DNA fragment was isolated and treated with calf intestinal phosphatase and then ligated with the 0.2kb *Bgl*III/*Bam*HI N terminal DX-890 cDNA to create pDB2683. The DNA and amino acid sequence of the N-terminal DX-890-(GGG)<sub>4</sub>GG-albumin fusion are shown in Table 14 and Table 15, respectively. Appropriate yeast vector sequences were provide by a "disintegration" plasmid pSAC35 generally disclosed in EP-A-286 424 and described by Sleep, D., *et al.* (1991) *Bio/Technology* 9, 183-187. The *Not*I N-terminal DX-890-(GGG)<sub>4</sub>GG-rHA expression cassette was isolated from pDB2683, purified and ligated into *Not*I digested pSAC35 which had been treated with calf intestinal phosphatase creating pDB2684 contained the *Not*I expression cassette in the opposite orientation to *LEU2*.

**Example 9: Construction of C-terminal albumin-(GGG)<sub>4</sub>GG-DX-890 expression plasmid**

Plasmid pDB2575X was partially digested with *Hind*III and then digested to completion with *Bam*HI. The desired 6.55kb DNA fragment was isolated and ligated with the 0.2kb *Bam*HI/*Hind*III C terminal DX-890 cDNA to create pDB2649. The DNA and amino acid sequence of the C-terminal albumin-(GGG)<sub>4</sub>GG-DX-890 fusion are shown in Table 16 and Table 17, respectively. Appropriate yeast vector sequences were provide by a "disintegration" plasmid pSAC35 generally disclosed in EP-A-286 424 and described by Sleep, D., *et al.* (1991) *Bio/Technology* 9, 183-187. The *Not*I C-terminal albumin-(GGG)<sub>4</sub>GG-DX-890 expression cassette was isolated from pDB2649, purified and ligated into

- 58 -

*NotI* digested pSAC35 which had been treated with calf intestinal phosphatase, creating two plasmids; the first pDB2652 contained the *NotI* expression cassette in the same expression orientation as *LEU2*, while the second pDB2653 contained the *NotI* expression cassette in the opposite orientation to *LEU2*.

#### **Example 10: Fermentation to Produce a Fusion Protein**

The DX-890-HSA fusion protein was expressed in fermentation culture as described in WO 00/44772. The DX-890-HSA fusion protein was purified from fermentation culture supernatant using the standard HA purification SP-FF (Pharmacia) conditions as described in WO 00/44772, except that an extra 200mM NaCl was required in the elution buffer.

#### **Example 11: Yeast transformation and culturing conditions**

Yeast strains disclosed in WO 95/23857, WO 95/33833 and WO 94/04687 were transformed to leucine prototrophy as described in Sleep D., *et al.* (2001) *Yeast* 18, 403-421. The transformants were patched out onto Buffered Minimal Medium (BMM, described by Kerry-Williams, S.M. *et al.* (1998) *Yeast* 14, 161-169) and incubated at 30 °C until grown sufficiently for further analysis.

#### **Example 12: $K_i$ Measurement of DX-890 Samples**

Equilibrium inhibition constants ( $K_i$ ) for DX-890 or DX-890-HSA inhibition of HNE were determined according to the tight-binding inhibition model with formation of a reversible complex (1:1 stoichiometry). Inhibition of hNE was determined at 30 °C in 50 mM HEPES, pH 7.5, 150 mM NaCl, and 0.1% Triton X-100. All reactions (total volume = 200  $\mu$ L) were carried out in microtiter plates (Costar #3789). hNE was incubated with varying concentrations of added inhibitor for 24 hours. Residual enzymatic activities were determined from the relative rates of substrate hydrolysis. The hydrolysis reaction was initiated by addition of N-methoxysuccinyl-Ala-Ala-Pro-Val-7-amino-methylcoumarin as substrate. Enzymatic cleavage of this substrate releases the methylcoumarin moiety with concomitant increase the sample fluorescence. The rate of substrate hydrolysis was monitored at an excitation of 360 nm and an emission of 460 nm. Plots of the percent remaining activity versus inhibitor concentration were fit by nonlinear regression analysis to Equation 1 to determine equilibrium dissociation constants.

$$\%A = 100 - \left( \frac{(I + E + K_i) - \sqrt{(I + E + K_i)^2 - 4 \cdot E \cdot I}}{2 \cdot E} \right) \cdot 100 \quad (1)$$

Where:

%A = percent activity

I = DX-890

E = HNE concentration

K<sub>i</sub> = equilibrium inhibition constant

The K<sub>i</sub> of native DX-890 was measured at the same time as a positive control. The K<sub>i</sub>'s of DX-890 and DX-890-HSA fusion for human neutrophil elastase (HNE) were similar to each other (Figure 1). Similar results were seen with the DX-890-HSA fusion in supernatant from a shake flask yeast culture or from a fermentor. Both supernatants were supplied by Aventis to Dyax. This result indicates that fusion to HSA does not affect the potency of DX-890 as an inhibitor of HNE.

### **Example 13: Fusions of DX-88 to N terminus of HSA**

DX-88 is a Kunitz domain derived from the first Kunitz domain of human LACI which inhibits human plasma kallikrein with K<sub>i</sub> ~ 40 pM. The serum half-time of DX-88 is not more than 1 hour. DX-88 is currently being tested in the clinic for treatment of hereditary angioedema (HAE). Initial data suggest that DX-88 is safe and effective. HAE is a condition in which attacks recur episodically and having a long-acting form would allow prophylactic treatment instead of reactive treatment.

A DNA sequence is available for DX-88, prepared for fusion to the N terminus of HA. The DNA sequences are provided at the 5' or 3' end to encode bridging sequences between the DX-88 coding region, the albumin coding region or the leader sequence as appropriate for N-terminal DX-88-(GGG)<sub>4</sub>GG-albumin (Table 18).

Plasmid pDB2573X is digested to completion with *Bgl*III and *Bam*HI, the 6.21kb DNA fragment is isolated and treated with calf intestinal phosphatase and then ligated with the 0.2kb *Bgl*III/*Bam*HI N terminal DX-88 cDNA to create pDB2666-88. The DNA and amino acid sequence of the N-terminal DX-88-(GGG)<sub>4</sub>GG-albumin fusion are shown in Table 19 and Table 20, respectively. Appropriate yeast vector sequences are provided by a "disintegration" plasmid pSAC35 generally disclosed in EP-A-286 424 and described by Sleep, D., *et al.* (1991) *Bio/Technology* 9, 183-187. The *Not*I N-terminal DX-88-(GGG)<sub>4</sub>GG-rHA expression

- 60 -

cassette is isolated from pDB2666-88, purified and ligated into *NotI* digested pSAC35 which had been treated with calf intestinal phosphatase, creating two plasmids; the first pDB2679-88 contains the *NotI* expression cassette in the same expression orientation as *LEU2*, while the second pDB2680-88 contains the *NotI* expression cassette in the opposite orientation to *LEU2*.

**Example 14: Construction of C-terminal albumin-(GGS)<sub>4</sub>GG-DX-88 expression plasmid**

As in Example 5, Plasmid pDB2575X is partially digested with *HindIII* and then digested to completion with *BamHI*. The desired 6.55kb DNA fragment is isolated and ligated with the 0.2kb *BamHI/HindIII* C terminal DX-88 cDNA (Table 21) to create pDB2648-88. The DNA and amino acid sequence of the C-terminal albumin-(GGS)<sub>4</sub>GG-DX-88 fusion are shown in Table 22 and Table 23, respectively. Appropriate yeast vector sequences are provided by a "disintegration" plasmid pSAC35 generally disclosed in EP-A-286 424 and described by Sleep, D., *et al.* (1991) *Bio/Technology* 9, 183-187. The *NotI* C-terminal albumin-(GGS)<sub>4</sub>GG-DX-88 expression cassette is isolated from pDB2648-88, purified and ligated into *NotI* digested pSAC35 which is treated with calf intestinal phosphatase, creating pDB2651-88 contained the *NotI* expression cassette in the same expression orientation as *LEU2*.

**Example 15: Pharmacokinetic Study in Mice**

The DX-890-HSA fusion protein was expressed in fermentation culture as described in WO 00/44772. The DX-890-HSA fusion protein was purified from fermentation culture supernatant using the standard HA purification SP-FF (Pharmacia) conditions as described in WO 00/44772, except that an extra 200mM NaCl was required in the elution buffer.

About 10 mg of rHA-DX-890 fusion was purified from the diafiltration retentate by SEC-HPLC and characterized by SCS-PAGE and RP-HPLC methods to be about 92% monomeric form. This material was used for subsequent <sup>125</sup>I radiolabeling and *in-vivo* plasma clearance studies.

For studies using mice, animals were injected in the tail vein and 4 animals were sacrificed at approximately 0, 7, 15, 30 and 90 minutes, 4h, 8h, 16h, 24h after injection, less 4 time points for the native DX-890 because of its likely short half life. Time of injection and time of sampling were recorded. At sacrifice, samples of ~0.5 ml were collected into anticoagulant (0.02 ml EDTA). Cells were spun down and separated from plasma. Plasma was divided into two aliquots, one frozen and one stored at 4 °C for immediate analysis. Analysis included gamma counting of all samples. In addition, analysis was performed for

- 61 -

two plasma samples (N=2) at each time point, i.e., 0, and 30 minutes, for  $^{125}\text{I}$ -DX-890, and 0, 30 minutes, and 24 h for the  $^{125}\text{I}$ -DX-890-HSA fusion. A SEC -HPLC Superose-12 column with an in-line radiation detector was used to analyze plasma fractions.

The results show that fusing DX-890 to HSA dramatically improves its beta (elimination) half life by ~5X (Figure 2). In addition, it appears that the DX-890-HSA-fusion is more stable in mouse plasma than DX-890 (Figures 3 and 4).

#### **Example 16: Pharmacokinetic Study in Rabbits**

Pharmacokinetic properties of DX-890 and DX-890-HSA were measured by iodinating the proteins and measuring clearance of the radiolabel from circulation in rabbits. The two DX-890 preparations were iodinated with iodine-125 using the iodogen method. After radiolabeling, the two labeled protein preparations were purified from unbound label by size exclusion chromatography (SEC). Fractions from the SEC column having the highest radioactivity were pooled. The purified, radiolabeled preparations were characterized for specific activity by scintillation counting and for purity by SEC using a Superose-12 column equipped with an in-line radiation detector.

New Zealand White rabbits (*ca.* 2.5 Kg) were used for clearance measurements, with one animal each used for of the two labeled protein preparations. The radiolabeled preparation was injected into the animal via an ear vein. One blood sample was collected per animal per time point with early time points at approximately 0, 7, 15, 30, and 90 minutes and later time points at 4, 8, 16, 24, 48, 72, 96, 144, 168, and 192 hours. Samples (about 0.5 ml) were collected into anticoagulant (EDTA) tubes. Cells were separated from the plasma/serum fraction by centrifugation. The plasma fraction was divided into two aliquots. One plasma aliquot was stored at  $-70^{\circ}\text{C}$  and the other aliquot was kept at  $4^{\circ}\text{C}$  for immediate analyses. Sample analyses included radiation counting for clearance rate determinations and SEC chromatography for *in vivo* stability. The results of the rabbit clearance study are summarized in Figures 5 and 6 and in Table 24.

The HSA-DX-890 fusion protein shows substantial improvements in *in vivo* circulation properties relative to those of the unmodified DX-890. Plasma clearance rates are greatly reduced for the fusion protein so that after a single day relative circulating levels of radiolabel are more than 100-fold higher for the HSA-DX-890 fusion than for the unmodified protein (Figure 5). A simple bi-exponential fit to the data shows large increases in both the alpha and beta portions of the clearance curve (Table 24). In particular, the value for  $T_{1/2\beta}$  is increased more than 20-fold, from about 165 min (2.75 hrs) for the unmodified protein to

- 62 -

about 3500 min (~ 60 hrs, ~ 2.5 days) for the HSA-DX-890 fusion. In addition, the fraction of the total material involved in the slow clearance portion of the curve nearly doubles for the fusion protein relative to unmodified DX-890 (Table 24).

**Table 24****Clearance Times in Rabbits**

Compound	Dose		Clearance Times (min)			
	$\mu\text{gm}$	$\mu\text{Ci}$	$T_{1/2\alpha}$	% $\alpha$	$T_{1/2\beta}$	% $\beta$
DX-890	50	83	0.4	75	165	25
HSA-DX-890	151	105	270	60	3500	40

Finally, *in vivo* stability appears to be improved for the fusion protein relative to unmodified DX-890 (Figure 6). SEC analysis of plasma from the rabbit injected with  $^{125}\text{I}$ -DX-890 (Figure 6, Part A) shows a relatively rapid association of label with higher molecular weight plasma components (earlier eluting peaks). Further, the relative proportion of the total residual circulating label associated with the high molecular weight material increases as time post-injection increases (compare 30 min and 4 hour elution profiles). In contrast, SEC analyses of plasma samples from the rabbit injected with  $^{125}\text{I}$ -HSA-DX-890 (Figure 6, Part B) shows that almost all of the circulating label is associated with the HSA-DX-890 peak seen in the injectate and that the label remains stably associated with this peak for at least 72 hours.

**Example 17: A Vector for Making a Doubly Fused HSA**

The vector pDB2300X1 is a modification of pDB2575X in which there is a *BglII/BamHI* cassette near the 5' terminus of the rHA gene and a *BspEI/KpnI* cassette near the 3' terminus. The *NotI* cassette that comprises this gene is shown in Table 25 showing the DNA, encoded AA sequence and useful restriction sites. In each line in Table 25, everything after an exclamation point is commentary, the DNA sequence is numbered and spaced to allow understand the design.

**Example 18: Adding a first instance of DX890 to pDB2300X1**

The DNA shown in Table 12 is introduced into pDB2300X1 that has been cut with *BglII* and *BamHI* to make the new vector pDB2300X2. The DNA, encoded AA sequence and useful restriction sites of the *NotI* cassette of pDB2300X2 are shown in Table 26.



- 63 -

**Example 19: Adding a second instance of DX890 to pDB2300X2**

The DNA shown in Table 27 is introduced into pDB2300X2 that has been cut with BspEI and KpnI to make the new vector pDB2300X3. Although this DNA encodes the same AA sequence as does the DNA of Table 12, many codons have been changed to reduce the likelihood of recombination between the two DX890-encoding regions. The DNA, encoded AA sequence and useful restriction sites of this construct are shown in Table 28. The encoded AA sequence is shown in Table 29. This protein is expressed in the same manner as the other constructions of the present invention. The protein of Table 103, "Dx890-HA-Dx890", will have ~ 16% the HNE-neutralizing activity of DX890 but a much long serum life time. Thus area-under-the-curve for inhibition of HNE will be much higher than for naked DX890.

**Example 20: DX1000::(GGS)4GG::HSA**

The DNA shown in Table 30 is introduced into pDB2573X which has been cut with BglII and BamHI to create pDX1000. The AA sequence of the encoded protein is shown in Table 31. Expression of this protein is essentially the same as for other HA fusions of the present invention.

**Example 21: DX-88::(GGS)4GG::HSA::(GGS)4GG::DX-88**

In a manner similar to the construction of a gene encoding DX-890-HSA-DX-890, the DNA of Table 18 is inserted into pDB2300X1 that has been cut with BglII and BamHI to make the new vector pDB2300X88a. The DNA shown in Table 32 is introduced into pDB2300X88a as a BspEI/KpnI fragment to create pDB2300X88b which contains two instances of DNA that encodes DX-88. The DNA in Table 32 is substantially different from the DNA in Table 18 so that recombination is unlikely.

**Example 22: Multiple Albumin Fusions**

The N-terminal fusion expression plasmid, pDB2540, as described herein, can be modified to introduce a unique Bsu36I at the C-terminal end; the new plasmid is named pDB2301X. The DNA sequence of the NotI expression cassette from pDB2301X is as follows:

pDB2540+Bsu36I

```

      NotI
1  GCGGCCGCcc gtaatgcggt atcgtgaaag cgaaaaaaaa actaacagta gataagacag
61 atagacagat agagatggac gagaaacagg gggggagaaa aggggaaaag agaaggaaaag

```

```

                                                NarI
121 aaagactcat ctatcgcaga taagacaatc aacctcatG GCGCCTccaa ccaccatccg

```

- 64 -

```

181 cactagggac caagcgctcg caccggttagc aacgcttgac tcacaaacca actgccggct
241 gaaagagctt gtgcaatggg agtgccaatt caaaggagcc gaatacgtct gtcgccttt
301 taagaggctt tttgaacact gcattgcacc cgacaaatca gccactaact acgaggtcac
361 ggacacatat accaatagtt aaaaattaca tatactctat atagcacagt agtgtgataa
421 ataaaaaatt ttgccaagac ttttttaaac tgcacccgac agatcaggtc tgtgcctact
481 atgcacttat gcccggggtc cggggaggag aaaaaacgag ggctgggaaa tgtccgtgga
541 ctttaaacgc tccgggttag cagagtagca gggctttcgg ctttggaat ttaggtgact
601 tgttgaaaaa gcaaaatttg ggctcagtaa tgccactgca gtggcttatc acgccaggac
661 tgcgggagtg gcgggggcaa acacaccgc gataaagagc gcgatgaata taaaagggg
721 ccaatgttac gtcccgttat attggagttc tccccatata aacttaagag tccaattagc

```

HindIII

```

781 ttcacgcga ataaaaaac AAGCTTaacc taattctaac aagcaaagat gaagtgggtt
>>.....>

```

BglII

```

841 ttcacgtct ccattttgtt cttgttctcc tctgcttact ctAGATCTtt ggataagaga
>.....Fusion Leader.....>

```

AgeI

```

901 gacgctcaca agtcogaagt cgctcACCGG Ttcaaggacc taggtgagga aaacttcaag
>>.....rHA synth. gene ..Continues to base 2655.....>
961 gcttttgtct tgatcgcttt cgctcaatac ttgcaacaat gtccattcga agatcacgtc
1021 aagtttgtca acgaagttac cgaattcgct aagacttggt ttgctgacga atctgctgaa
1081 aactgtgaca agtccttgca cacttggttc ggtgataagt tgtgtactgt tgctaccttg
1141 agagaaacct acggtgaaat ggctgactgt tgtgctaagc aagaaccaga aagaaacgaa
1201 tgtttcttgc aacacaagga cgacaacca aacttgccaa gattgggttag accagaagtt
1261 gacgtcatgt gtactgcttt ccacgacaac gaagaaacct tcttgaagaa gtacttgtac
1321 gaaattgcta gaagacaccc atacttctac gctccagaat tgttgttctt gcctaagaga
1381 tacaaggctg ctttcaccga atgttgtcaa gctgctgata aggtgcttg tttgttgcca
1441 aagttggatg aattgagaga cgaaggtaag gcttcttcg ctaagcaaag attgaagtgt
1501 gcttcttgc aaaagttcgg tgaaagagct ttcaaggctt gggctgtcgc tagattgtct
1561 caaagattcc caaaggctga attcgctgaa gtttctaagt tgggtactga cttgactaag
1621 gttcacactg aatgttgtca cggtgacttg ttggaatgtg ctgatgacag agctgacttg
1681 gctaagtaca tctgtgaaaa ccaagactct atctcttcca agttgaagga atgttgtgaa
1741 aagccattgt tggaaaagtc tcaactgtatt gctgaagttg aaaacgatga aatgccagct
1801 gacttgccat ctttggtgc tgacttcgtt gaacttaagg acgtttgtaa gaactacgt
1861 gaagctaagg acgtcttctt gggatgttct ttgtacgaat acgctagaag acaccagac
1921 tactccgttg tcttgtgtt gagattggct aagacctacg aaactacctt ggaaaagtgt
1981 tgtgctgctg ctgaccaca cgaatgttac gctaaggttt tcgatgaatt caagccattg
2041 gtcgaagaac cacaaaactt gatcaagcaa aactgtgaat tgttcgaaca attgggtgaa
2101 tacaagttcc aaaacgcttt gttgggtaga tacactaaga aggtcccaca agtctccacc
2161 ccaactttgg ttgaagtctc tagaaacttg ggtaaggctg gttctaagtg ttgtaagcac
2221 ccagaagcta agagaatgcc atgtgctgaa gattacttgt ccgtcgtttt gaaccaattg
2281 tgtgttttgc acgaaaagac cccagtctct gatagagtca ccaagtgttg tactgaatct
2341 ttggttaaca gaagaccatg tttctctgct ttggaagtcg acgaaactta cgttccaaag

```

EcoRV

```

2401 gaattcaacg ctgaaacttt cacttccac gctGATATct gtaccttgct cgaaggagaa
2461 agacaaatta agaagcaaac tgctttgggt gaattgggtca agcacaagcc aaagggtact
2521 aaggaacaat tgaaggctgt catggatgat ttcgctgctt tcgttgaaaa gtgttgtaag
2581 gctgatgata aggaaacttg tttcgctgaa gaaggtaaga agttggctgc tgcttcccaa

```

Bsu36I

HindIII

```

2641 gctgCCTTAG GcttataatA AGCTTaattc ttatgattta tgatttttat tattaataaa
>>.....>
2701 gttataaaaa aaataagtgt atacaaattt taaagtgact cttagggtttt aaaacgaaaa
2761 ttcttattct tgagtaactc tttctgttag gtcagggtgc tttctcaggt atagcatgag

```

SphI

```

2821 gtcgctctta ttgaccacac ctctaccgGC ATGCcgagca aatgcctgca aatcgctccc

```

- 65 -

2881 catttcaccc aattgtagat atgctaactc cagcaatgag ttgatgaatc tcggtgtgta

NotI

2941 ttttatgtcc tcagaggaca acacctgttg taatcggtct tccacacgga tcGCGGCCGC

DNA encoding polypeptides can be inserted in between the *Bgl*II and *Age*I sites to express an N-terminal albumin fusion, or between the *Bsu*36I and *Hind*III (not unique and so will require a partial *Hind*III digest) sites to express an C-terminal albumin fusion, or between both pairs of sites to make a co-N- and C-terminal albumin fusion.

Polypeptide spacers can be optionally incorporated. The DNA sequence of the *Not*I expression cassette from the modified pDB2540 is expected to be as follows:

pDB2540+2xGSlinkers

NotI

1 GCGGCCGCcc gtaatgcggt atcgtgaaag cgaaaaaaaa actaacagta gataagacag  
61 atagacagat agagatggac gagaaacagg gggggagaaa aggggaaaaag agaaggaaaag

NarI

121 aaagactcat ctatcgaga taagacaatc aaccctcatG GCGCCTccaa ccaccatccg  
181 cactagggac caagcgctcg caccgttagc aacgcttgac tcacaaacca actgccggct  
241 gaaagagctt gtgcaatggg agtgccaatt caaaggagcc gaatacgtct gctgccttt  
301 taagaggctt tttgaacact gcattgcacc cgacaaatca gccactaact acgaggtcac  
361 ggacacatat accaatagtt aaaaattaca tatactctat atagcacagt agtgtgataa  
421 ataaaaaatt ttgccaagac ttttttaaac tgcacccgac agatcaggtc tgtgcctact  
481 atgcacttat gcccggggtc ccgggaggag aaaaaacgag ggctgggaaa tgtccgtgga  
541 ctttaaacgc tccgggttag cagagtagca gggctttcgg ctttggaat ttaggtgact  
601 tggtgaaaaa gcaaaatttg ggctcagtaa tgccactgca gtggcttatc acgccaggac  
661 tgcgggagtg gcgggggcaa acacaccgc gataaagagc gcgatgaata taaaagggg  
721 ccaatgttac gtcccggtat attggagttc ttccataca aacttaagag tccaattagc

HindIII

781 ttcacgcca ataaaaaaaa AAGCTTaacc taattctaac aagcaaagat gaagtgggtt  
>>.....>

BglII

841 ttcacgtctt ccattttgtt cttgttctcc tctgcttact ctAGATCTtt ggataagaga  
>.....Fusion Leader.....>>

BamHI

901 ggtGGATCCg gtggttcggt tgggtctggt ggttcgggtg gtgacgctca caagtccgaa  
>>.....GS linker.....>|>>.....rHA.....>

AgeI

961 gtcgctcACC GGTtcaagga cctaggtgag gaaaacttca aggctttggt cttgatcgct  
>.....rHA synth. gene continues to base 2739.....>

1021 ttcgctcaat acttgcaaca atgtccattc gaagatcacg tcaagtgtgt caacgaagtt  
1081 accgaattcg ctaagacttg tgttgctgac gaatctgctg aaaactgtga caagtccttg  
1141 cacaccttgt tcggtgataa gttgtgtaact gttgctacct tgagagaaac ctacggtgaa  
1201 atggctgact gttgtgctaa gcaagaacca gaaagaaaac aatgtttctt gcaacacaag  
1261 gacgacaacc caaacttgcc aagattgggt agaccagaag ttgacgtcat gtgtactgct  
1321 ttccacgaca acgaagaaac cttcttgaag aagtacttgt acgaaattgc tagaagacac  
1381 ccatacttct acgctccaga attgttggtc ttcgctaaga gatacaaggc tgctttcacc

- 66 -

```

1441  gaatgttgtc aagctgctga taaggctgct tgtttgttgc caaagttgga tgaattgaga
1501  gacgaaggta aggtctcttc cgctaagcaa agattgaagt gtgcttcctt gcaaaagttc
1561  ggtgaaagag ctttcaaggc ttgggctgtc gctagattgt ctcaaagatt cccaaaggct
1621  gaattcgctg aagtttctaa gttggttact gacttgacta aggttcacac tgaatgttgt
1681  cacggtgact tgttggaatg tgctgatgac agagctgact tggctaagta catctgtgaa
1741  aaccaagact ctatctcttc caagttgaag gaatgttgtg aaaagccatt gttggaaaag
1801  tctcactgta ttgctgaagt tgaaaacgat gaaatgccag ctgacttgcc atctttggct
1861  gctgacttcg ttgaatctaa ggacgtttgt aagaactacg ctgaagctaa ggacgtcttc
1921  ttgggtatgt tcttgtagca atacgctaga agacacccag actactccgt tgtcttgttg
1981  ttgagattgg ctaagaccta cgaaactacc ttggaaaagt gttgtgctgc tgctgacca
2041  cacgaatgtt acgctaagg tttcgatgaa ttcaagccat tggctgaaga accacaaaac
2101  ttgatcaagc aaaactgtga attgttcgaa caattgggtg aatacaagtt ccaaaacgct
2161  ttgttggtta gatacactaa gaaggctcca caagtctcca cccaacttt ggttgaagtc
2221  tctagaaact tgggtaaggc cggttctaag tgttgtaagc acccagaagc taagagaatg
2281  ccatgtgctg aagattactt gtcctgctgt ttgaaccaat tgtgtgtttt gcacgaaaag
2341  accccagtct ctgatagagt caccaagtgt tgtactgaat ctttggttaa cagaagacca
2401  tgtttctctg ctttggaagt cgacgaaact tacgttccaa aggaattcaa cgctgaaact

```

EcoRV

```

2461  ttcaccttcc acgctGATAT Ctgtaccttg tccgaaaagg aaagacaaat taagaagcaa
2521  actgctttgg ttgaattggt caagcacaa ccaaaggcta ctaaggaaca attgaaggct
2581  gtcattgatg atttcgctgc tttcgttgaa aagtgttgta aggtgatga taaggaaact

```

Bsu36I

```

2641  tgtttcgctg aagaaggtaa gaagttggtc gctgcttccc aagctgCCTT AGGcttaggt
>.....rHA synth. gene .....>|>>>

```

BspEI

KpnI

HindIII

```

2701  ggttctggtg gtTCCGGAgg ttctggtGGT ACCggtggtt aatAAGCTTa attcttatga
>.....GS linker.....>>

```

```

2761  tttatgattt ttattattaa ataagttata aaaaaaataa gtgtatacaa attttaaagt
2821  gactcttagg ttttaaaacg aaaattctta ttcttgagta actctttcct gtaggtcagg

```

SphI

```

2881  ttgctttctc aggtatagca tgaggtcgtc cttattgacc acacctctac cgGCATGCcg
2941  agcaaatgcc tgcaaatcgc tccccatttc acccaattgt agatatgcta actccagcaa
3001  tgagttgatg aatctcggtg tgtattttat gtcctcagag gacaacacct gttgtaatcg

```

NotI

```

3061  ttcttccaca cggatcGCGG CCGC

```

DNA encoding polypeptides can be inserted in between the *Bg/III* and *BamHI* sites to express an N-terminal albumin fusion, or between the unique BspEI and KpnI sites to express an C-terminal albumin fusion, or between both pairs of sites to make a co-N- and C-terminal albumin fusion. This is exemplified most simply by using the *Bg/III-BamHI* DPI-14 cDNA and the *BamHI-HindIII* DX-890 cDNA as described herein. By ligating these cDNAs into the appropriate site, a DPI-14-(GGG)<sub>4</sub>GG-rHA-(GGG)<sub>4</sub>GG-DX-890 fusion with the following DNA sequence would be constructed.

NotI

```

1   GCGGCCGCcc gtaatgcggt atcgtgaaag cgaaaaaaaa actaacagta gataagacag
61  atagacagat agagatggac gagaaacagg gggggagaaa aggggaaaag agaaggaaaag

```

- 67 -

## NarI

```

121 aaagactcat ctatcgaga taagacaatc aaccctcatG GCGCCTccaa ccaccatccg
181 cactagggac caagcgctcg caccgttagc aacgcttgac tcacaaacca actgccggct
241 gaaagagctt gtgcaatggg agtgccaatt caaaggagcc gaatacgtct gctcgccctt
301 taagaggctt tttgaacact gcattgcacc cgacaaatca gccactaact acgaggtcac
361 ggacacatat accaatagtt aaaaattaca tatactctat atagcacagt agtgtgataa
421 ataaaaaatt ttgccaagac ttttttaaac tgcacccgac agatcaggct tgtgcctact
481 atgcacttat gcccggggtc ccgggaggag aaaaaacgag ggctgggaaa tgtccgtgga
541 ctttaaagcg tccgggttag cagagtagca gggctttcgg ctttggaat ttaggtgact
601 tggtgaaaaa gcaaaatttg ggctcagtaa tgccactgca gtggcttatc acgccaggac
661 tgcgggagtg gcgggggcaa acacaccgc gataaagagc gcgatgaata taaaagggg
721 ccaatgttac gtcccgttat attggagttc ttcccatata aacttaagag tccaattagc

```

## HindIII

```

781 ttcacgcca ataaaaaac AAGCTTaacc taattctaac aagcaaagat gaagtgggtt
>>.....>

```

## BglII

```

841 ttcacgtct ccattttgtt cttgttctcc tctgcttact ctAGATCTtt ggataagaga
>.....Fusion Leader.....>>

901 gaagctgtta gagaagtttg ttctgaacaa gctgaaactg gtccatgtat tgctttcttc
>>.....DPI-14 up to base 1080.....>

961 ccaagatggg acttcgatgt tactgaaggt aagtgcgcgc cattcttcta cgggtggtgt
1021 ggtggaataa gaaacaactt cgatactgaa gaatactgta tggtgtttg tggttctgct
>.....DPI-14.....>>

```

## BamHI

```

1081 ggtGGATCCg gtgggtccgg tggttctggg ggttccgggtg gtgacgtca caagtcggaa
>>.....GS linker.....>|>>...rHA synth gene.>

```

## AgeI

```

1141 gtcgctcACC GGTtcaagga ctaggtgag gaaaacttca aggctttggg cttgatcgct
>.....rHA synth. gene continues to base 2877.....>

```

```

1201 ttcgctcaat acttgcaaca atgtccattc gaagatcacg tcaagttggg caacgaagtt
1261 accgaattcg ctaagacttg tggtgctgac gaatctgctg aaaactgtga caagtccttg
1321 cacaccttgt tcggtgataa gttgtgtact gttgctacct tgagagaaac ctacggtgaa
1381 atggctgact gttgtgctaa gcaagaacca gaaagaaacg aatgtttctt gcaacacaag
1441 gacgacaacc caaacttgcc aagattgggt agaccagaag ttgacgtcat gtgtactgct
1501 ttccacgaca acgaagaaac cttcttgaag aagtacttgt acgaaattgc tagaagacac
1561 ccatacttct acgctccaga attgttggtc ttcgctaaga gatacaaggc tgctttcacc
1621 gaatgttggtc aagctgctga taaggctgct tgtttgttg caaagttgga tgaattgaga
1681 gacgaaggta aggcttcttc cgctaagcaa agattgaagt gtgcttcctt gcaaaagttc
1741 ggtgaaagag ctttcaaggc ttgggctgtc gctagattgt ctcaaagatt cccaaaggct
1801 gaattcgctg aagtttctaa gttgggtact gacttgacta aggttcacac tgaatgttgt
1861 cacgggtgact tggtggaatg tgctgatgac agagctgact tggctaagta catctgtgaa
1921 aaccaagact ctatctcttc caagttgaag gaatgttgtg aaaagccatt gttggaaaag
1981 tctcactgta ttgctgaagt tgaaaacgat gaaatgccag ctgacttgcc atctttggct
2041 gctgacttcg ttgaatctaa ggacgtttgt aagaactacg ctgaagctaa ggacgtcttc
2101 ttgggtatgt tcttgtagca atacgctaga agacaccag actactccgt tgtcttggtg
2161 ttgagattgg ctaagaccta cgaactacc ttggaaaagt gttgtgctgc tgctgaccca
2221 cacgaatggt acgctaaggt tttcgtgaa ttcaagccat tgggtcgaaga accacaaaac
2281 ttgatcaagc aaaactgtga attgttcgaa caattgggtg aatacaagtt ccaaaacgct
2341 ttgttggtta gataactaa gaaggtccca caagtctcca ccccaacttt ggttgaagtc
2401 tctagaaact tgggtaaggt cggttctaa tggttgaagc acccagaagc taagagaatg
2461 ccatgtgctg aagattactt gtccgtcggt ttgaaccaat tgtgtgtttt gcacgaaaag
2521 accccagtct ctgatagagt caccaagtgt tgtactgaat ctttggttaa cagaagacca
2581 tgtttctctg ctttggaagt cgacgaaact tacgttccaa aggaattcaa cgctgaaact
2641 ttcaccttcc acgctGATAT CTgtaccttg tccgaaaagg aaagacaaat taagaagcaa

```

- 68 -

```

2701 actgcttttg ttgaattggt caagcacaag ccaaaggcta ctaaggaaca attgaaggct
2761 gtcattggtg atttcgctgc ttctggtgaa aagtgttgta aggctgatga taaggaaact

                                     Bsu36I
2821 tgtttcgctg aagaaggtaa gaagttggtc gctgcttccc aagctgCCTT AGGcttaggt
>.....rHA synth. gene .....>|>>>

                                     BspEI
2881 ggttctggtg gtTCCGGAgg tagtggtggc tccggtggtg aggcttgcaa tcttcctatc
Linker----->|--DX-890 (second coding)-->

2941 gtccgtggcc cttgcatcgc cttttttcct cgttgggcct ttgacgccgt caaaggcaaa
3001 tgcgtccttt ttccttacgg cggttgccag ggcaatggca ataaatttta tagcgagaaa
3061 gagtgcctg agtattgcgg cgtcccttaa taaGGTACct aatAAGCTTa attcctatga
----DX-890 (2nd coding)---->|

3121 tttatgattt ttattattaa ataagttata aaaaaataa gtgtatacaa attttaaagt
3181 gactcttagg ttttaaaacg aaaattctta ttcttgagta actctttcct gtaggtcagg

                                     SphI
3241 ttgctttctc aggtatagca tgaggctcgt cttattgacc acacctctac cgGCATGCcg
3301 agcaaatgcc tgcaaatcgc tccccatttc acccaattgt agatatgcta actccagcaa
3361 tgagttgatg aatctcggtg tgtattttat gtctcagag gacaacacct gttgtaatcg

                                     NotI
3421 ttcttcaca cggatcGCGG CCGC

```

The primary translation product of this DPI-14-(GGG)<sub>4</sub>GG-rHA-(GGG)<sub>4</sub>GG-DX-890 fusion is as follows.

```

1  MKWVFIVSIL FLFSSAYSRS LDKREAVREV CSEQAETGPC IAFFPRWYFD
51  VTEGKCAPFF YGGCGGNRNN FDTEEYCMAY CGSAGGSGGS GSGGSGGGDA
101 HKSEVAHRFK DLGEENFKAL VLIAFAQYLQ QCPFEDHVKL VNEVTEFAKT
151 CVADESAENC DKSLHTLFGD KLCTVATLRE TYGEMADCCA KQEPERNECF
201 LQHKDDNPNL PRLVRPEVDV MCTAFHDNEE TFLKKYLYEI ARRHPYFYAP
251 ELLFFAKRYK AAFTECCQAA DKAACLLPKL DELRDEGKAS SAKQRLKCAS
301 LQKFGERAFK AWAVARLSQR FPKAEFAEVS KLVTDLTKVH TECCHGDLLE
351 CADDRADLAK YICENQDSIS SKLKECCEKP LLEKSHCIAE VENDEMPADL
401 PSLAADFVES KDVCKNYAEA KDVFLGMFLY EYARRHPDYS VVLLRLAKT
451 YETTLEKCCA AADPHECYAK VFDEFKPLVE EPQNLIKQNC ELFEQLGEYK
501 FQNALLVRYT KKVPQVSTPT LVEVSRNLGK VGSKCCKHPE AKRMPCAEDY
551 LSVVLNQLCV LHEKTPVSDR VTKCTESLV NRRPCFSALE VDETYVPKEF
601 NAETFTFHAD ICTLSEKERQ IKKQTALVEL VKHKPKATKE QLKAVMDDFA
651 AFVEKCKKAD DKETCFAEEG KKLVAASQAA LGLGGSGGSG GSGGSGGEAC
701 NLPIVRGPCI AFFPRWAFDA VKGKCVLFPY GGCQGNNGKF YSEKECREYC
751  GVP

```

But as the first 24 amino acids constitute the fusion leader sequence, as described herein, the amino acid sequence of the secreted product are as follows:

```

1  EAVREVCSEQ AETGPCIAFF PRWYFDVTEG KCAPFFYGGC GGNRNNFDTE
51  EYCMAYCGSA GSGGSGGSGS GSGGDAHKSE VAHRFKDLGE ENFKALVLIA

```

- 69 -

```

101  FAQYLQQCPF EDHVKLVNEV TEFAKTCVAD ' ESAENCDSKL HTLFGDKLCT
151  VATLRETYGE MADCCAKQEP ERNECFLQHK DDNPNLPLRV RPEVDVMCTA
201  FHDNEETFLK KYLYEIARRH PYFYAPELLF FAKRYKAAFT ECCQAADKAA
251  CLLPKLDEL R DEGKASSAKQ RLKCASLQKF GERAFAKAWAV ARLSQRFPKA
301  EFAEVSKLVT DLTKVHTECC HGDLLLECADD RADLAKYICE NQDSISSKLK
351  ECCEKPLLEK SHCIAEVEND EMPADLPSLA ADFVESKDVC KNYAEAKDVF
401  LGMFLYEYAR RHPDYSVLL LRLAKTYETT LEKCCAAADP HECYAKVFDE
451  FKPLVEEPQN LIKQNCLEFE QLGEYKFQNA LLVRYTKKVP QVSTPTLVEV
501  SRNLGKVGSK CCKHPEAKRM PCAEDYLSVV LNQLCVLHEK TPVSDRVTKC
551  CTESLVNRRP CFSALEVDET YVPKEFNAET FTFHADICTL SEKERQIKKQ
601  TALVELVKHK PKATKEQLKA VMDDFAAFVE KCKKADDKET CFAEEGKKLV
651  AASQAALGLG GSGGSGGSGG SGGEACNLPI VRGPCIAFFP RWAFFDAVKKG
701  CVLFPYGGCQ GNGNKFYSEK ECREYCGVP

```

### **EXAMPLE 23: Amino-Acid Sequence of a DPI-14-(GGG)<sub>4</sub>GG-HSA Fusion Protein**

Table 33 shows the amino-acid sequence of a fusion of DPI14 via a linker comprising (GGG)<sub>4</sub>GG to HSA. Construction of a gene to encode the given sequence is simple using the methods and vectors described herein. DPI-14 is a potent inhibitor of HNE and the fusion to HSA produces a molecule with longer serum residence time.

### **Tables:**

**Table 1: Amino-acid sequencer of Mature HSA from GenBank entry AAN17825**

```

DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV KLVNEVTEFA
KTCVADESAE NCDKSLHTLF GDKLCTVATL RETYGE MADC CAKQEPERNE
CFLQHKDDNP NLPRLVRPEV DVMCTAFHDN EETFLKKYLY EIARRHPYFY
APELLFFAKR YKAAFTTECCQ AADKAACLLP KLDEL RDEGK ASSAKQRLKC
ASLQKFGERA FKAWARELS QRFPKAEFAE VSKLVTDLTK VHTECHGDL
LECADDRADL AKYICENQDS ISSKLKECCE KPLLEKSHCI AEVENDEMPA
DLPSLAADFV ESKDVCKNYA EAKDVFLGMF LYEYARRHPD YSVLLLLRLA
KTYKTTLEKC CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE
YKFQNAL LVR YTKKVPQVST PTLVEVSRNL GKVGSCKCKH PEAKRMPCAE
DYLSVVLNQL CVLHEKTPVS DRVTKCCTES LVNRRPCFSA LEVDETYVPK
EFNAETFTFH ADICTLSEKE RQIKKQ TALV ELVKHKPKAT KEQLKAVMDD
FAAFVEKCKK ADDKETCF AE EGKKLV AASR AALGL (SEQ ID NO:18)

```

- 70 -

**Table 2: Amino-acid sequences of DX-1000 and DX-88****DX-1000**

EAMHSFCAFKAETGPCRARFDRWFFNIFTRQCEEFIYGGCEGNQNRFSLEECKKMCTRD  
(SEQ ID NO: \_\_\_\_)

**DX-88**

EAMHSFCAFKADDGPCRAAHPRWFFNIFTRQCEEFIYGGCEGNQNRFSLEECKKMCTRD  
(SEQ ID NO: \_\_\_\_)

**Table 5: DNA sequence of the N-terminal *Bgl*III-*Bam*HI DPI-14 cDNA**

AGATCTTTGGATAAGAGAGAAGCTGTTAGAGAAGTTTGTCTGAACAAGCTGAAACTGGTCCAT  
GTATTGCTTTCTTCCCAAGATGGTACTTCGATGTTACTGAAGGTAAGTGCGCGCCATTCTTCTA  
CGGTGGTTGTGGTGGTAACAGAAACAACTTCGATACTGAAGAATACTGTATGGCTGTTTGTGGT  
TCTGCTGGTGGATCC (SEQ ID NO: \_\_\_\_)

**Table 6: DNA sequence of the C-terminal *Bam*HI-*Hind*III DPI-14 cDNA**

GGATCCGGTGGTGAAGCTGTTAGAGAAGTTTGTCTGAACAAGCTGAAACTGGTCCATGTATTG  
CTTTCTTCCCAAGATGGTACTTCGATGTTACTGAAGGTAAGTGCGCGCCATTCTTCTACGGTGG  
TTGTGGTGGTAACAGAAACAACTTCGATACTGAAGAATACTGTATGGCTGTTTGTGGTTCTGCT  
TAATAAGCTT (SEQ ID NO: \_\_\_\_)

**Table 7: DNA sequence of the N-terminal  
DPI-14-(GGS)<sub>4</sub>GG-albumin fusion coding region**

GAAGCTGTTAGAGAAGTTTGTCTGAACAAGCTGAAACTGGTCCATGTATTGCTTTCTTCCCAA  
GATGGTACTTCGATGTTACTGAAGGTAAGTGCGCGCCATTCTTCTACGGTGGTTGTGGTGGTAA  
CAGAAACAACTTCGATACTGAAGAATACTGTATGGCTGTTTGTGGTTCTGCTGGTGGATCCGGT  
GGTTCCGGTGGTTCTGGTGGTTCCGGTGGTGACGCTCACAAGTCCGAAGTCGCTCACC GGTTCA  
AGGACCTAGGTGAGGAAAACCTTCAAGGCTTTGGTCTTGATCGCTTTCGCTCAATACTTGCAACA  
ATGTCCATTGCAAGATCACGTCAAGTTGGTCAACGAAGTTACCGAATTCGCTAAGACTTGTGTT  
GCTGACGAATCTGCTGAAAACCTGTGACAAGTCCTTGACACCTTGTTCCGGTGATAAGTTGTGTA  
CTGTTGCTACCTTGAGAGAAACCTACGGTGAAATGGCTGACTGTTGTGCTAAGCAAGAACCAGA  
AAGAAACGAATGTTTCTTGCAACACAAGGACGACAACCCAACTTGCCAAGATTGGTTAGACCA  
GAAGTTGACGTCATGTGTACTGCTTTCCACGACAACGAAGAAACCTTCTTGAAGAAGTACTTGT  
ACGAAATTGCTAGAAGACACCCATACTTCTACGCTCCAGAATTGTTGTTCTTCGCTAAGAGATA  
CAAGGCTGCTTTCACCGAATGTTGTCAAGCTGCTGATAAGGCTGCTTGTTTGTGTTGCCAAAGTTG  
GATGAATTGAGAGACGAAGGTAAGGCTTCTTCCGCTAAGCAAAGATTGAAGTGTGCTTCCTTGC



- 71 -

AAAAGTTCGGTGAAAGAGCTTTCAAGGCTTGGGCTGTCTGCTAGATTGTCTCAAAGATTCCCAAA  
 GGCTGAATTCGCTGAAGTTTCTAAGTTGGTTACTGACTTGACTAAGGTTACACTGAATGTTGT  
 CACGGTGACTTGTGGAATGTGCTGATGACAGAGCTGACTTGGCTAAGTACATCTGTGAAAACC  
 AAGACTCTATCTCTTCCAAGTTGAAGGAATGTTGTGAAAAGCCATTGTTGGAAAAGTCTCACTG  
 TATTGCTGAAGTTGAAAACGATGAAATGCCAGCTGACTTGCCATCTTTGGCTGCTGACTTCGTT  
 GAATCTAAGGACGTTTGTGAAGAACTACGCTGAAGCTAAGGACGTCTTCTTGGGTATGTTCTTGT  
 ACGAATACGCTAGAAGACACCCAGACTACTCCGTTGTCTTGTGTTGAGATTGGCTAAGACCTA  
 CGAAACTACCTTGAAAAGTGTGTTGTGCTGCTGCTGACCCACACGAATGTTACGCTAAGGTTTTTC  
 GATGAATTCAAGCCATTGGTCGAAGAACCACAAAACCTTGATCAAGCAAACTGTGAATTGTTTCG  
 AACAAATTGGGTGAATACAAGTTCCAAAACGCTTGTGTTGGTTAGATACACTAAGAAGGTCCCACA  
 AGTCTCCACCCCAACTTTGGTTGAAGTCTCTAGAACTTGGGTAAGGTCGGTTCTAAGTGTGTT  
 AAGCACCCAGAAGCTAAGAGAATGCCATGTGCTGAAGATTACTTGTCCGTCGTTTTGAACCAAT  
 TGTGTGTTTTTGCACGAAAAGACCCAGTCTCTGATAGAGTCACCAAGTGTGTTACTGAATCTTT  
 GGTTAACAGAAGACCATGTTTTCTCTGCTTTGGAAGTCGACGAACTTACGTTCCAAAGGAATTC  
 AACGCTGAAACTTTCACCTTCCACGCTGATATCTGTACCTTGTCCGAAAAGGAAAGACAAATTA  
 AGAAGCAAACTGCTTTGGTTGAATTGGTCAAGCACAAGCCAAAGGCTACTAAGGAACAATTGAA  
 GGCTGTCATGGATGATTTCGCTGCTTTGCTTGAAAAGTGTGTTAAGGCTGATGATAAGGAACT  
 TGTTTCGCTGAAGAAGGTAAGAAGTTGGTCGCTGCTTCCCAAGCTGCTTTGGGTTTG (SEQ  
 ID NO: \_\_\_\_ )

**Table 8: Amino acid sequence of the N-terminal  
DPI-14-(GGS)<sub>4</sub>GG-albumin fusion protein**

EAVREVCSEQAETGPCIAFFPRWYFDVTEGKCAPFFYGGCGGNRNNFDTEEYCMVCGSAGGSG  
 GSGSGSGSGGDAHKSEVAHRFKDLGEENFKALVLI AFAQYLQQCPFEDHVKLVNEVTEFAKTCV  
 ADESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQE PERNECFLQHKDDNP NLPRLVRP  
 EVDVMCTAFHDNEETFLKKYLYE IARRHPYFYAPELLFFAKRYKAAFTECCQAADKAACLLPKL  
 DELRDEGKASSAKQRLKCASLQKFGERAFKAWAVARLSQRFPKAEFAEVSKLVTDLT KVHTECC  
 HGDLL ECADDRADLAKYI CENQDS ISSKLKECCEKPLLEKSHCIAEVENDEMPADLPSLAADFV  
 ESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVVL LRLAKTYETTLEKCCAAADPHECYAKVF  
 DEFKPLVEEPQNLIKQNC ELFELGEYKFQNAL LVRYTKKVPQVSTPTLVEVSRNLGKVGSKCC  
 KHPEAKRMPCAEDYLSVVLNQLCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPKEF  
 NAETFTFHADI CTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCKKADDKET  
 CFAEEGKKLV AASQAALGL (SEQ ID NO: \_\_\_\_ )

- 72 -

**Table 9: DNA sequence of the C-terminal  
albumin-(GGG)<sub>4</sub>GG-DPI-14 fusion coding region**

GATGCACACAAGAGTGAGGTTGCTCATCGGTTTAAAGATTTGGGAGAAGAAAATTTCAAAGCCT  
TGGTGGTTGATTGCCTTTGCTCAGTATCTTCAGCAGTGTCCATTTGAAGATCATGTAAAATTAGT  
GAATGAAGTAACTGAATTTGCAAAAACATGTGTTGCTGATGAGTCAGCTGAAAATTGTGACAAA  
TCACTTCATACCCCTTTTTGGAGACAAATTATGCACAGTTGCAACTCTTCGTGAAACCTATGGTG  
AAATGGCTGACTGCTGTGCAAAACAAGAACCTGAGAGAAATGAATGCTTCTTGCAACACAAAGA  
TGACAACCCAAACCTCCCCGATTGGTGAGACCAGAGGTTGATGTGATGTGCACTGCTTTTCAT  
GACAATGAAGAGACATTTTTGAAAAATACTTATATGAAATTGCCAGAAGACATCCTTACTTTT  
ATGCCCCGGAACCTCCTTTTCTTTGCTAAAAGGTATAAAGCTGCTTTTACAGAATGTTGCCAAGC  
TGCTGATAAAGCTGCCTGCCTGTTGCCAAAGCTCGATGAACCTTCGGGATGAAGGGAAGGCTTCG  
TCTGCCAAACAGAGACTCAAGTGTGCCAGTCTCCAAAAATTTGGAGAAAGAGCTTTCAAAGCAT  
GGGCAGTAGCTCGCCTGAGCCAGAGATTTCCCAAAGCTGAGTTTGCAGAAGTTTCCAAGTTAGT  
GACAGATCTTACCAAAGTCCACACGGAATGCTGCCATGGAGATCTGCTTGAATGTGCTGATGAC  
AGGGCGGACCTTGCCAAGTATATCTGTGAAAATCAAGATTCGATCTCCAGTAAACTGAAGGAAT  
GCTGTGAAAACCTCTGTTGGAAAAATCCCACTGCATTGCCGAAGTGGAATGATGAGATGCC  
TGCTGACTTGCCCTCATTAGCTGCTGATTTTGTGAAAGTAAGGATGTTTGCAAAAACCTATGCT  
GAGGCAAAGGATGTCTTCCTGGGCATGTTTTTGTATGAATATGCAAGAAGGCATCCTGATTACT  
CTGTGCTGCTGCTGCTGAGACTTGCCAAGACATATGAAACCACTCTAGAGAAGTGCTGTGCCGC  
TGCAGATCCTCATGAATGCTATGCCAAAGTGTTTCGATGAATTTAAACCTCTTGTGGAAGAGCCT  
CAGAATTTAATCAAACAAAATTGTGAGCTTTTTTGAGCAGCTTGAGAGTACAAATTCAGAATG  
CGCTATTAGTTCGTTACACCAAGAAAGTACCCCAAGTGTCAACTCCAACCTCTTGTAGAGGTCTC  
AAGAAACCTAGGAAAAGTGGGCAGCAAATGTTGTAAACATCCTGAAGCAAAAAGAATGCCCTGT  
GCAGAAGACTATCTATCCGTGGTCCTGAACCAGTTATGTGTGTTGCATGAGAAAACGCCAGTAA  
GTGACAGAGTCACCAAATGCTGCACAGAATCCTTGGTGAACAGGCGACCATGCTTTTCAGCTCT  
GGAAGTCGATGAAACATACGTTCCCAAAGAGTTTAATGCTGAAACATTACCTTCCATGCAGAT  
ATATGCACACTTTCTGAGAAGGAGAGACAAATCAAGAAACAACTGCACTTGTTGAGCTCGTGA  
AACACAAGCCCAAGGCAACAAAAGAGCAACTGAAAGCTGTTATGGATGATTTTCGCAGCTTTTGT  
AGAGAAGTGCTGCAAGGCTGACGATAAGGAGACCTGCTTTGCCGAGGAGGGTAAAAAACTTGTT  
GCTGCAAGTCAAGCTGCCTTAGGCTTAGGTGGTTCTGGTGGTTCCGGTGGTTCTGGTGGATCCG  
GTGGTGAAGCTGTTAGAGAAGTTTGTCTGAACAAGCTGAACTGGTCCATGTATTGCTTTCTT  
CCCAAGATGGTACTTCGATGTTACTGAAGGTAAGTGCGCGCCATTCTTCTACGGTGGTTGTGGT  
GGTAACAGAAACAACTTCGATACTGAAGAATACTGTATGGCTGTTTGTGGTTCTGCT (SEQ

ID NO: \_\_\_\_)

- 73 -

**Table 10: Amino acid sequence of the C-terminal albumin-(GGG)<sub>4</sub>GG-DPI-14 fusion protein**

DAHKSEVAHRFKDLGEENFKALVLI AFAQYLQQCPFEDHVKLVNEVTEFAKTCVADESAENCDK  
 SLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNLPRLVRPEVDVMCTAFH  
 DNEETFLKKYLYEIIARRHPYFYAPELLFFAKRYKAAFTECCQAADKAACLLPKLDEL RDEGKAS  
 SAKQRLKCASLQKFGERAFAKAWAVARLSQRFPAEFAEVSKLVTDLT KVHTECCHGDLLECADD  
 RADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPADLPSLAADFVESKDVCKNYA  
 EAKDVFLGMFLYEYARRHPDYSVLLLLRLAKTYETTTLEKCCAAADPHECYAKVFDEFKPLVEEP  
 QNLIKQNC ELF EQLGEYKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCKKHPEAKRMPC  
 AEDYLSVVLNQLCVLHEKTPVSDRVTKCTESLVNRRPCFSALEVD ETYVPKEFNAETFTFHAD  
 ICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCKADDKETCF AE EGKKLV  
 AASQAALGLGGSGGSGGSGGSGGEAVREVCSEQAETGPCIAFFPRWYFDVTEGKCAPFFYGGCG  
 GNRNNFDTEEYCMAVCGSA (SEQ ID NO: \_\_\_\_)

**Table 11: DNA sequence of the C-terminal BamHI-HindIII DX-1000 cDNA**

GGA TCC GGT GGT  
 gag gct atg cat tcc ttc tgc gcc ttc aag  
 gct gag act ggt cct tgt aga gct agg ttc  
 gac cgt tgg ttc ttc aac atc ttc acg cgt  
 cag tgc gag gaa ttc att tac ggt ggt tgt  
 gaa ggt aac cag aac cgg ttc gaa tct cta  
 gag gaa tgt aag aag atg tgc act cgt gac  
 TAA TAA GCT T (SEQ ID NO: \_\_\_\_)

**Table 12: DNA sequence of the N-terminal BglIII-BamHI DX-890 cDNA**

AGATCTTTGGATAAGAGAGAAGCCTGTAACCTTGCCAATTGTTAGAGGTCCATGTATTGCTTTCT  
 TCCCAAGATGGGCTTTTCGATGCTGTTAAGGGTAAGTGTGTTTTGTTCCCATATGGTGGTTGTCA  
 AGGTAACGGTAACAAGTTCTACTCTGAAAAGGAATGTAGAGAATACTGTGGTGTTCAGGTGGA  
 TCC (SEQ ID NO: \_\_\_\_)

**Table 13: DNA sequence of the C-terminal BamHI-HindIII DX-890 cDNA**

GGATCCGGTGGTGAAGCCTGTAACCTTGCCAATTGTTAGAGGTCCATGTATTGCTTTCTTCCCAA  
 GATGGGCTTTTCGATGCTGTTAAGGGTAAGTGTGTTTTGTTCCCATATGGTGGTTGTCAAGGTAA  
 CGGTAACAAGTTCTACTCTGAAAAGGAATGTAGAGAATACTGTGGTGTTCATAATAAGCTT  
 (SEQ ID NO: \_\_\_\_)

- 74 -

**Table 14: DNA sequence of the N-terminal  
DX-890-(GGS)<sub>4</sub>GG-albumin fusion coding region**

GAAGCCTGTAACCTTGCCAATTGTTAGAGGTCCATGTATTGCTTTCTTCCCAAGATGGGCTTTTCG  
ATGCTGTTAAGGGTAAGTGTGTTTTGTTCCCATATGGTGGTTGTCAAGGTAACGGTAACAAGTT  
CTACTCTGAAAAGGAATGTAGAGAATACTGTGGTGTTCAGGTGGATCCGGTGGTTCCGGTGGT  
TCTGGTGGTTCCGGTGGTGACGCTCACAAGTCCGAAGTCGCTCACCGGTTCAAGGACCTAGGTG  
AGGAAAACCTTCAAGGCTTTGGTCTTGATCGCTTTCGCTCAATACTTGCAACAATGTCCATTCTGA  
AGATCACGTCAAGTTGGTCAACGAAGTTACCGAATTTCGCTAAGACTTGTGTTGCTGACGAATCT  
GCTGAAAACCTGTGACAAGTCCTTGACACCTTGTTCGGTGATAAGTTGTGTACTGTTGCTACCT  
TGAGAGAAACCTACGGTGAAATGGCTGACTGTTGTGCTAAGCAAGAACCAGAAAGAAACGAATG  
TTTCTTGCAACACAAGGACGACAACCCAAACTTGCCAAGATTGGTTAGACCAGAAGTTGACGTC  
ATGTGTACTGCTTTCACGACAACGAAGAAACCTTCTTGAAGAAGTACTTGTACGAAATTGCTA  
GAAGACACCCATACTTCTACGCTCCAGAATTGTTGTTCTTCGCTAAGAGATACAAGGCTGCTTT  
CACCGAATGTTGTCAAGCTGCTGATAAGGCTGCTTGTGTTGTTGCCAAAGTTGGATGAATTGAGA  
GACGAAGGTAAGGCTTCTTCCGCTAAGCAAAGATTGAAGTGTGCTTCCTTGCAAAAGTTCCGGTG  
AAAGAGCTTTCAAGGCTTGGGCTGTGCTAGATTGTCTCAAAGATTCCCAAAGGCTGAATTTCGC  
TGAAGTTTCTAAGTTGGTTACTGACTTGACTAAGGTTCACTGAATGTTGTCACGGTGACTTG  
TTGGAATGTGCTGATGACAGAGCTGACTTGCTAAGTACATCTGTGAAAACCAAGACTCTATCT  
CTTCCAAGTTGAAGGAATGTTGTGAAAAGCCATTGTTGAAAAGTCTCACTGTATTGCTGAAGT  
TGAAAACGATGAAATGCCAGCTGACTTGCCATCTTTGGCTGCTGACTTCGTTGAATCTAAGGAC  
GTTTGTAAAGAACTACGCTGAAGCTAAGGACGTCTTCTTGGGTATGTTCTTGTACGAATACGCTA  
GAAGACACCCAGACTACTCCGTTGTCTTGTGTTGAGATTGGCTAAGACCTACGAACTACCTT  
GGAAAAGTGTTGTGCTGCTGCTGACCCACACGAATGTTACGCTAAGGTTTTTCGATGAATTCAAG  
CCATTGGTCGAAGAACCACAAAACCTTGATCAAGCAAACTGTGAATTGTTTCAACAATTGGGTG  
AATACAAGTTCCAAAACGCTTTGTTGGTTAGATACACTAAGAAGGTCCACAAAGTCTCCACCCC  
AACTTTGGTTGAAGTCTCTAGAACTTGGGTAAGGTCGGTTCTAAGTGTGTAAGCACCAGAA  
GCTAAGAGAATGCCATGTGCTGAAGATTACTTGTCCGTCGTTTTGAACCAATTGTGTGTTTTGC  
ACGAAAAGACCCAGTCTCTGATAGAGTCACCAAGTGTGTACTGAATCTTTGGTTAACAGAAG  
ACCATGTTTCTCTGCTTTGGAAGTCGACGAACTTACGTTCCAAAGGAATTCAACGCTGAACT  
TTCACCTTCCACGCTGATATCTGTACCTTGTCCGAAAAGGAAAGACAAATTAAGAAGCAAACCTG  
CTTTGGTTGAATTGGTCAAGCACAAGCCAAAGGCTACTAAGGAACAATTGAAGGCTGTCATGGA  
TGATTTTCGCTGCTTTCGTTGAAAAGTGTTGTAAGGCTGATGATAAGGAACTTGTTCGCTGAA  
GAAGGTAAGAAGTTGGTCGCTGCTTCCCAAGCTGCTTTGGGTTTG (SEQ ID NO: \_\_\_\_)

- 75 -

**Table 15: Amino acid sequence of the N-terminal  
DX-890-(GGS)<sub>4</sub>GG-albumin fusion protein**

EACNLP IVRGPCIAFFPRWAFDAVKGKCVLFPYGGCQGNNGNFYSEKECREYCGVPGSGSGSGG  
 SGGSGGDAHKSEVAHRFKDLGEENFKALVLI AFAQYLQQCPFEDHV KLVNEVTEFAKTCVADES  
 AENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQE PERNECF LQHKDDNP NLPRLV RPEVDV  
 MCTAFHDNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAA FTECCQAADKAA CLLPKLDEL R  
 DEGKASSAKQRLK CASLQKFGERAFKAWAVARLSQRFPKAEFAEVSKLVTDLT KVHTECCHGDL  
 LECADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPADLPSLAADFVESKD  
 VCKNYAEAKDVFLGMFLY EYARRHPDYSV VLLLRLAKTYETTLEKCCAAADPHECYAKVFDEFK  
 PLVEEPQNL I KQNC ELF EQLGEYKFQNAL LVRYTKKVPQVSTPTLVEVSRNLGKV GSKCKHPE  
 AKRMPCAEDYLSVVLNQLCVLHEKTPVSDRVTKCTESLVNRRPCFSALEVD ETYVPKEFNAET  
 FTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCKADDKETCF AE  
 EGKKLVAASQAALGL (SEQ ID NO: \_\_\_\_)

**Table 16: DNA sequence of the C-terminal  
albumin-(GGS)<sub>4</sub>GG-DX-890 fusion coding region**

GATGCACACA AGAGTGAGGT TGCTCATCGG TTAAAGATT TGGGAGAAGA  
 AAATTTCAAA GCCTTGGTGT TGATTGCCTT TGCTCAGTAT CTTCAGCAGT  
 GTCCATTTGA AGATCATGTA AAATTAGTGA ATGAAGTAAC TGAATTTGCA  
 AAAACATGTG TTGCTGATGA GTCAGCTGAA AATTGTGACA AATCACTTCA  
 TACCCTTTTT GGAGACAAAT TATGCACAGT TGCAACTCTT CGTGAAACCT  
 ATGGTGAAAT GGCTGACTGC TGTGCAAAAC AAGAACCTGA GAGAAATGAA  
 TGCTTCTTGC AACACAAAGA TGACAACCCA AACCTCCCCC GATTGGTGAG  
 ACCAGAGGTT GATGTGATGT GCACTGCTTT TCATGACAAT GAAGAGACAT  
 TTTTGAAAAA ATA CT TATAT GAAATTGCCA GAAGACATCC TTACTTTTAT  
 GCCCCGGAAC TCCTTTTCTT TGCTAAAAGG TATAAAGCTG CTTTTACAGA  
 ATGTTGCCAA GCTGCTGATA AAGCTGCCTG CCTGTTGCCA AAGCTCGATG  
 AACTTCGGGA TGAAGGGAAG GCTTCGTCTG CCAAACAGAG ACTCAAGTGT  
 GCCAGTCTCC AAAAATTTGG AGAAAGAGCT TTCAAAGCAT GGGCAGTAGC  
 TCGCCTGAGC CAGAGATTTT CCAAAGCTGA GTTTGCAGAA GTTTCCAAGT  
 TAGTGACAGA TCTTACCAA GTCCACACGG AATGCTGCCA TGGAGATCTG  
 CTTGAATGTG CTGATGACAG GGCGGACCTT GCCAAGTATA TCTGTGAAAA  
 TCAAGATTCG ATCTCCAGTA AACTGAAGGA ATGCTGTGAA AAACCTCTGT  
 TGGAAAAATC CCACTGCATT GCCGAAGTGG AAAATGATGA GATGCCTGCT  
 GACTTGCCTT CATTAGCTGC TGATTTTGTT GAAAGTAAGG ATGTTTGCAA

- 76 -

AAACATGCT GAGGCAAAGG ATGTCTTCCT GGGCATGTTT TTGTATGAAT  
 ATGCAAGAAG GCATCCTGAT TACTCTGTCTG TGCTGCTGCT GAGACTTGCC  
 AAGACATATG AAACCACTCT AGAGAAGTGC TGTGCCGCTG CAGATCCTCA  
 TGAATGCTAT GCCAAAGTGT TCGATGAATT TAAACCTCTT GTGGAAGAGC  
 CTCAGAATTT AATCAAACAA AATTGTGAGC TTTTGTGAGCA GCTTGGAGAG  
 TACAAATTCC AGAATGCGCT ATTAGTTCGT TACACCAAGA AAGTACCCCA  
 AGTGTCAACT CCAACTCTTG TAGAGGTCTC AAGAAACCTA GGAAAAGTGG  
 GCAGCAAATG TTGTAAACAT CCTGAAGCAA AAAGAATGCC CTGTGCAGAA  
 GACTATCTAT CCGTGGTCCT GAACCAGTTA TGTGTGTTGC ATGAGAAAAC  
 GCCAGTAAGT GACAGAGTCA CCAATGCTG CACAGAATCC TTGGTGAACA  
 GGCGACCATG CTTTTTCAGCT CTGGAAGTCG ATGAAACATA CGTTCCCAAA  
 GAGTTTAAATG CTGAAACATT CACCTTCCAT GCAGATATAT GCACACTTTC  
 TGAGAAGGAG AGACAAATCA AGAAACAAAC TGCACCTGTT GAGCTCGTGA  
 AACACAAGCC CAAGGCAACA AAAGAGCAAC TGAAAGCTGT TATGGATGAT  
 TTCGCAGCTT TTGTAGAGAA GTGCTGCAAG GCTGACGATA AGGAGACCTG  
 CTTTGCCGAG GAGGGTAAAA AACTTGTTGC TGCAAGTCAA GCTGCCTTAG  
 GCTTAGGTGG TTCTGGTGGT TCCGGTGGTT CTGGTGGATC CGGTGGTGAA  
 GCCTGTAACT TGCCAATTGT TAGAGGTCCA TGTATTGCTT TCTTCCAAG  
 ATGGGCTTTC GATGCTGTTA AGGGTAAGTG TGTTTTGTTT CCATATGGTG  
 GTTGTCAAGG TAACGGTAAC AAGTTCTACT CTGAAAAGGA ATGTAGAGAA  
 TACTGTGGTG TTCCA (SEQ ID NO: \_\_\_\_ )

**Table 17 Amino acid sequence of the C-terminal  
albumin-(GGS)<sub>4</sub>GG-DX-890 fusion protein**

DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV KLVNEVTEFA KTCVADESAE  
 NCDKSLHTLF GDKLCTVATL RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV  
 DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR YKAAFTECCQ AADKAACLLP  
 KLDEL RDEGK ASSAKQRLKC ASLQKFGERA FKAWAVARLS QRFPAEFAE VSKLVTDLTK  
 VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE KPLLEKSHCI AEVENDEMPA  
 DLPSLAADFV ESKDVCKNYA EAKDVFLGMF LYEYARRHPD YSVVLLLRLLA KTYETTLEKC  
 CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE YKFQNALIVR YTKKVPQVST  
 PTLVEVSRNL GKVGSKCKKH PEAKRMPCAE DYLSVVLNQL CVLHEKTPVS DRVTCKCTES  
 LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEKE RQIKKQTALV ELVKHKPKAT  
 KEQLKAVMDD FAAFVEKCK ADDKETCF AE EGKKLVAAASQ AALGLGGSGG SGGSGGSGGE  
 ACNLPVIRGP CIAFFPRWAF DAVKGKCVLF PYGGCQGN GN KFYSEKECREY CGVP  
 (SEQ ID NO: \_\_\_\_ )

- 77 -

**Table 18: DNA sequence of the N-terminal *Bgl*III-*Bam*HI DX-88 cDNA**

AGA TCT TTG GAT AAG AGA  
 GAA GCT ATG CAC  
 TCT TTC TGT GCT TTC AAG GCT GAC GAC GGT  
 CCG TGC AGA GCT GCT CAC CCA AGA TGG TTC  
 TTC AAC ATC TTC ACG CGA CAA TGC GAG GAG  
 TTC ATC TAC GGT GGT TGT GAG GGT AAC CAA  
 AAC AGA TTC GAG TCT CTA GAG GAG TGT AAG  
 AAG ATG TGT ACT AGA GAC GGT GGA TCC (SEQ ID NO: \_\_\_\_)

**Table 19: DNA sequence of the N-terminal  
DX-88-(GGS)<sub>4</sub>GG-albumin fusion coding region**

GAA GCT ATG CAC TCT TTC TGT GCT TTC AAG GCT GAC GAC GGT CCG  
 TGC AGA GCT GCT CAC CCA AGA TGG TTC TTC AAC ATC TTC ACG CGA  
 CAA TGC GAG GAG TTC ATC TAC GGT GGT TGT GAG GGT AAC CAA AAC  
 AGA TTC GAG TCT CTA GAG GAG TGT AAG AAG ATG TGT ACT AGA GAC GGT  
 GGATCC  
 GGTGGTTCCGGTGGTTCTGGTGGTTCCGGTGGTGACGCTCACAAGTCCGAAGTCGCTCACCGGT  
 TCAAGGACCTAGGTGAGGAAACTTCAAGGCTTTGGTCTTGATCGCTTTCGCTCAATACTTGCA  
 ACAATGTCCATTCTGAAGATCACGTCAAGTTGGTCAACGAAGTTACCGAATTCGCTAAGACTTGT  
 GTTGCTGACGAATCTGCTGAAACTGTGACAAGTCCTTGACACACCTTGTTTCGGTGATAAGTTGT  
 GTACTGTTGCTACCTTGAGAGAAACCTACGGTGAAATGGCTGACTGTTGTGCTAAGCAAGAACC  
 AGAAAGAAACGAATGTTTCTTGCAACACAAGGACGACAACCCAAACTTGCCAAGATTGGTTAGA  
 CCAGAAGTTGACGTCATGTGTACTGCTTTCACGACAACGAAGAAACCTTCTTGAAGAAGTACT  
 TGTACGAAATTGCTAGAAGACACCCATACTTCTACGCTCCAGAATTGTTGTTCTTCGCTAAGAG  
 ATACAAGGCTGCTTTCACCGAATGTTGTCAAGCTGCTGATAAGGCTGCTTGTTTGTGTTGCCAAAG  
 TTGGATGAATTGAGAGACGAAGGTAAGGCTTCTTCCGCTAAGCAAAGATTGAAGTGTGCTTCCT  
 TGCAAAAGTTTCGGTGAAAGAGCTTTCAGGCTTGGGCTGTCGCTAGATTGTCTCAAAGATTCCC  
 AAAGGCTGAATTCGCTGAAGTTTCTAAGTTGGTACTGACTTGAAGTTTCACTGAATGT  
 TGTACGGTGACTTGTTGGAATGTGCTGATGACAGAGCTGACTTGGCTAAGTACATCTGTGAAA  
 ACCAAGACTCTATCTCTTCCAAGTTGAAGGAATGTTGTGAAAAGCCATTGTTGAAAAGTCTCA  
 CTGTATTGCTGAAGTTGAAAACGATGAAATGCCAGCTGACTTGCCATCTTTGGCTGCTGACTTC  
 GTTGAATCTAAGGACGTTTGTAAAGAACTACGCTGAAGCTAAGGACGTCTTCTTGGGTATGTTCT  
 TGTACGAATACGCTAGAAGACACCCAGACTACTCCGTTGTCTTGTGTTGAGATTGGCTAAGAC  
 CTACGAAACTACCTTGGAAGTGTGTGCTGCTGCTGACCCACACGAATGTTACGCTAAGGTT

- 78 -

TTCGATGAATTCAAGCCATTGGTCGAAGAACCACAAAACCTTGATCAAGCAAACTGTGAATTGT  
 TCGAACAATTGGGTGAATACAAGTTCCAAAACGCTTTGTTGGTTAGATACACTAAGAAGGTCCC  
 ACAAGTCTCCACCCCAACTTTGGTTGAAGTCTCTAGAACTTGGGTAAGGTCGGTTCTAAGTGT  
 TGTAAGCACCCAGAAGCTAAGAGAATGCCATGTGCTGAAGATTACTTGTCCGTCGTTTTGAACC  
 AATTGTGTGTTTTTGCACGAAAAGACCCAGTCTCTGATAGAGTCACCAAGTGTTGTACTGAATC  
 TTTGGTTAACAGAAGACCATGTTTCTCTGCTTTGGAAGTCGACGAACTTACGTTCCAAAGGAA  
 TTCAACGCTGAACTTTTACCTTCCACGCTGATATCTGTACCTTGTCGAAAAGGAAAGACAAA  
 TTAAGAAGCAAAGTCTTTGGTTGAATTGGTCAAGCACAAGCCAAAGGCTACTAAGGAACAATT  
 GAAGGCTGTCATGGATGATTTTCGCTGCTTTTCGTTGAAAAGTGTTGTAAGGCTGATGATAAGGAA  
 ACTTGTTCGCTGAAGAAGGTAAGAAGTTGGTCGCTGCTTCCCAAGCTGCTTTGGGTTTG

(SEQ ID NO: \_\_\_\_\_)

**Table 20: AA sequence of DX-88::HSA**

EAMHSFCAFK ADDGPCRAAH PRWFFNIFTR QCEEFIYGGC EGNQNRFESL  
 EECKKMCTRD GSGGSGSGSG GSGGDAHKSE VAHRFKDLGE ENFKALVLIA  
 FAQYLQQCPF EDHVKLNEV TEFAKTCVAD ESAENCCKSL HTLFGDKLCT  
 VATLRETYGE MADCCAKQEP ERNECFLOHK DDNPNLPLRV RPEVDVMCTA  
 FHDNEETFLK KYLYEIARRH PYFYAPELLE FAKRYKAAFT ECCQAADKAA  
 CLLPKLDELRL DEGKASSAKQ RLKCASLQKF GERAFAKAWAV ARLSQRFPA  
 EFAEVSKLVT DLTKVHTECC HGDLLCADD RADLAKYICE NQDSISSKLK  
 ECCEKPLLEK SHCIAEVEND EMPADLPSLA ADFVESKDVC KNYAEAKDVF  
 LGMFLYEYAR RHPDYSVLL LRLAKTYETT LEKCCAAADP HECYAKVFDE  
 FKPLVEEPQN LIQNCLEFE QLGEYKFQNA LLVRYTKKVP QVSTPTLVEV  
 SRNLGKVGSK CCKHPEAKRM PCAEDYLSVV LNQLCVLHEK TPVSDRVTKC  
 CTESLVNRRP CFSALEVDET YVPKEFNAET FTFHADICTL SEKERQIKKQ  
 TALVELVKHK PKATKEH (SEQ ID NO: \_\_\_\_\_)

**Table 21: DNA sequence of the C-terminal BamHI-HindIII DX-88 cDNA**

GGA TCC GGT GGT GAA GCT ATG CAC  
 TCT TTC TGT GCT TTC AAG GCT GAC GAC GGT  
 CCG TGC AGA GCT GCT CAC CCA AGA TGG TTC  
 TTC AAC ATC TTC ACG CGA CAA TGC GAG GAG  
 TTC ATC TAC GGT GGT TGT GAG GGT AAC CAA  
 AAC AGA TTC GAG TCT CTA GAG GAG TGT AAG  
 AAG ATG TGT ACT AGA GAC



- 79 -

TAA TAA GCT T (SEQ ID NO: \_\_\_\_\_)

**Table 22: HSA::(GGS)4GG::DX-88**

gat gca cac aag agt gag gtt gct cat cgg ttt aaa gat ttg gga  
gaa gaa aat ttc aaa gcc ttg gtg ttg att gcc ttt gct cag tat  
ctt cag cag tgt cca ttt gaa gat cat gta aaa tta gtg aat gaa  
gta act gaa ttt gca aaa aca tgt gtt gct gat gag tca gct gaa  
aat tgt gac aaa tca ctt cat acc ctt ttt gga gac aaa tta tgc  
aca gtt gca act ctt cgt gaa acc tat ggt gaa atg gct gac tgc  
tgt gca aaa caa gaa cct gag aga aat gaa tgc ttc ttg caa cac  
aaa gat gac aac cca aac ctc ccc cga ttg gtg aga cca gag gtt  
gat gtg atg tgc act gct ttt cat gac aat gaa gag aca ttt ttg  
aaa aaa tac tta tat gaa att gcc aga aga cat cct tac ttt tat  
gcc ccg gaa ctc ctt ttc ttt gct aaa agg tat aaa gct gct ttt  
aca gaa tgt tgc caa gct gct gat aaa gct gcc tgc ctg ttg cca  
aag ctc gat gaa ctt cgg gat gaa ggg aag gct tcg tct gcc aaa  
cag aga ctc aag tgt gcc agt ctc caa aaa ttt gga gaa aga gct  
ttc aaa gca tgg gca gta gct cgc ctg agc cag aga ttt ccc aaa  
gct gag ttt gca gaa gtt tcc aag tta gtg aca gat ctt acc aaa  
gtc cac acg gaa tgc tgc cat gga gat ctg ctt gaa tgt gct gat  
gac agg gcg gac ctt gcc aag tat atc tgt gaa aat caa gat tcg  
atc tcc agt aaa ctg aag gaa tgc tgt gaa aaa cct ctg ttg gaa  
aaa tcc cac tgc att gcc gaa gtg gaa aat gat gag atg cct gct  
gac ttg cct tca tta gct gct gat ttt gtt gaa agt aag gat gtt  
tgc aaa aac tat gct gag gca aag gat gtc ttc ctg ggc atg ttt  
ttg tat gaa tat gca aga agg cat cct gat tac tct gtc gtg ctg  
ctg ctg aga ctt gcc aag aca tat gaa acc act cta gag aag tgc  
tgt gcc gct gca gat cct cat gaa tgc tat gcc aaa gtg ttc gat  
gaa ttt aaa cct ctt gtg gaa gag cct cag aat tta atc aaa caa  
aat tgt gag ctt ttt gag cag ctt gga gag tac aaa ttc cag aat  
gcg cta tta gtt cgt tac acc aag aaa gta ccc caa gtg tca act  
cca act ctt gta gag gtc tca aga aac cta gga aaa gtg ggc agc  
aaa tgt tgt aaa cat cct gaa gca aaa aga atg ccc tgt gca gaa  
gac tat cta tcc gtg gtc ctg aac cag tta tgt gtg ttg cat gag

- 80 -

aaa acg cca gta agt gac aga gtc acc aaa tgc tgc aca gaa tcc  
 ttg gtg aac agg cga cca tgc ttt tca gct ctg gaa gtc gat gaa  
 aca tac gtt ccc aaa gag ttt aat gct gaa aca ttc acc ttc cat  
 gca gat ata tgc aca ctt tct gag aag gag aga caa atc aag aaa  
 caa act gca ctt gtt gag ctc gtg aaa cac aag ccc aag gca aca  
 aaa gag caa ctg aaa gct gtt atg gat gat ttc gca gct ttt gta  
 gag aag tgc tgc aag gct gac gat aag gag acc tgc ttt gcc gag  
 gag ggt aaa aaa ctt gtt gct gca agt caa gct gcc tta ggc tta  
 ggt ggt tct ggt ggt tcc ggt ggt tct ggt gga tcc ggt ggt  
 GAA GCT ATG CAC TCT TTC TGT GCT TTC AAG GCT GAC GAC GGT CCG  
 TGC AGA GCT GCT CAC CCA AGA TGG TTC TTC AAC ATC TTC ACG CGA  
 CAA TGC GAG GAG TTC ATC TAC GGT GGT TGT GAG GGT AAC CAA AAC  
 AGA TTC GAG TCT CTA GAG GAG TGT AAG AAG ATG TGT ACT AGA GAC  
 (SEQ ID NO: \_\_\_\_\_)

**Table 23: AA sequence of mature protein encoded in Table 22**

DAHKSEVAHRFKDLGEENFKALVLI AFAQY  
 LQQCPFEDHVKLVNEVTEFAKTCVADESAE  
 NCDKSLHTLFGDKLCTVATLRETYGEMADC  
 CAKQEPERNECFLOHKDDNP NLPRLVRPEV  
 DVMCTAFHDNEETFLKKYLYEIARRHPYFY  
 APELLFFAKRYKAAFTECCQAADKAACLLP  
 KLDEL RDEGKASSAKQRLKASLQKFGERA  
 FKAWAVARLSQRFPKAEFAEVSKLVTDLT K  
 VHTECCHGDLLECADDRADLAKYICENQDS  
 ISSKLKECCEKPLLEKSHCIAEVENDEMPA  
 DLPSLAADFVESKDVCKNYAEAKDVFLGMF  
 LYEYARRHPDYSVVL LRLAKTYETTLEKC  
 CAAADPHECYAKVFDEFKPLVEEPQNLIKQ  
 NCELFEQLGEYKFQNAL LVRYTKKVPQVST  
 PTLVEVSRNLGKVGSKCKHPEAKRMPCAE  
 DYLSVVLNQLCVLHEKTPVSDRVTKCTES  
 LVNRRPCFSALEVDETYVPKEFNAETFTFH  
 ADICTLSEKERQIKKQTALVELVKHKPKAT

- 81 -

KEQLKAVMDDFAAFVEKCKADDKETCFAE  
 EGKKLVAASQAALGLGGSGGSGGSGGSGGE  
 AMHSFCAFKADDGPCRAAHPRWFFNIFTRQ  
 CEEFIYGGCEGNQNRFSLEECKKMCTRD  
 (SEQ. ID NO: \_\_\_\_\_)

**Table 25: NotI cassette of pDB2300X1 with 2xGS linkers**

```

!
!   1  GCGGCCGCcc gtaatgcggt atcgtgaaag cgaaaaaaaa actaacagta gataagacag
!   NotI....
!
!   61  atagacagat agagatggac gagaaacagg gggggagaaa aggggaaaag agaaggaaaag
121  aaagactcat ctatcgcaga taagacaatc aaccctcatG GCGCCTccaa ccaccatccg
!                                     NarI...
!
!   181 cactagggac caAGCGCTcg caccgttagc aacgcttgac tcacaaacca actGCCGGCT
!                                     AfeI..                               NgoMIV
!
!   241 gaaagagctt gtgcaatggg agtgccaatt caaaggagcc gaatacgtct gctcgccttt
301  taagaggctt tttgaacact gcattgcacc cgacaaatca gccactaact acgagggtcac
361  ggacacatat accaatagtt aaaaattaca tatactctat atagcacagt agtgtgataa
421  ataaaaaatt ttgccaagac ttttttaaaC TGCACccgac agatcagggtc tgtgcctact
!                                     BsgI...
!
!   481 atgcacttat gcccggggtc ccgggaggag aaaaaacgag ggctgggaaa tgtccgtgga
541  ctttaaacgc tccgggttag cagagtaGCA gggcttTCGg ctttggaaat ttaggtgact
!                                     BcgI.....
!
!   601 tgttgaaaaa gcaaaatttg ggctcagtaa tgCCAActgca gTGGcttatc acgccaggac
!                                     BstXI.....
!                                     PStI...
!
!   661 tgcggggagtg gcgggggcaa acacaccgcg gataaagagc gcgatgaata taaaaggggg
721  ccaatgttac gtcccgttat attggagttc ttcccataca aaCTTAAGag tccaattagc
!                                     AflII.
!
!   781 ttcatcgcca ataaaaaaaaac AAGCTTaacc taattctaac aagcaaag
!                                     HindIII (1/2)
!
!   1   2   3   4   5   6   7   8   9   10  11  12  13  14  15
!   M   K   W   V   F   I   V   S   I   L   F   L   F   S   S
!   829 atg aag tgg gtt ttc atc gtc tcc att ttg ttc ttg ttc tcc tct
!
!   16  17  18  19  20  21  22  23  24  25  26  27  28  29  30
!   A   Y   S   R   S   L   D   K   R   G   G   S   G   G   S
!   874 gct tac tct AGA TCT ttg gat aag aga ggt GGA TCC ggt ggt tcc
!                                     BglII..                               BamHI..
!
!   31  32  33  34  35  36  37  38  39  40  41  42  43  44  45
!   G   G   S   G   G   S   G   G   D   A   H   K   S   E   V
!   919 ggt ggt tct ggt ggt tcc ggt ggt gac gct cac aag tcc gaa gtc
!
!   46  47  48  49  50  51  52  53  54  55  56  57  58  59  60
!   A   H   R   F   K   D   L   G   E   E   N   F   K   A   L
!   964 gct cAC CGG Ttc aag gaC CTA GGt gag gaa aac ttc aag gct ttg
!                                     AgeI....                               AvrII...

```

- 82 -

```

!
!      61  62  63  64  65  66  67  68  69  70  71  72  73  74  75
!      V   L   I   A   F   A   Q   Y   L   Q   Q   C   P   F   E
! 1009  gtc ttg atc gct ttc gct caa tac ttg caa caa tgt cca ttc gaa
!
!      76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
!      D   H   V   K   L   V   N   E   V   T   E   F   A   K   T
! 1054  gat CAC GTC aag ttg gtc aac gaa gtt acc gaa ttc gct aag act
!      BmgBI..
!
!      91  92  93  94  95  96  97  98  99 100 101 102 103 104 105
!      C   V   A   D   E   S   A   E   N   C   D   K   S   L   H
! 1099  tgt gtt gct gac gaa tct gct gaa aac tgt gac aag tcc ttg cac
!
!      106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
!      T   L   F   G   D   K   L   C   T   V   A   T   L   R   E
! 1144  acc ttg ttc ggt gat aag ttg tgt act gtt gct acc ttg aga gaa
!
!      121 122 123 124 125 126 127 128 129 130 131 132 133 134 135
!      T   Y   G   E   M   A   D   C   C   A   K   Q   E   P   E
! 1189  acc tac ggt gaa atg gct gac tgt tgt gct aag caa gaa cca gaa
!
!      136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
!      R   N   E   C   F   L   Q   H   K   D   D   N   P   N   L
! 1234  aga aac gaa tgt ttc ttg caa cac aag gac gac aac cca aac ttg
!
!      151 152 153 154 155 156 157 158 159 160 161 162 163 164 165
!      P   R   L   V   R   P   E   V   D   V   M   C   T   A   F
! 1279  cca aga ttg gtt aga cca gaa gtt gac gtc atg tgt act gct ttc
!
!      166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
!      H   D   N   E   E   T   F   L   K   K   Y   L   Y   E   I
! 1324  cac gac aac gaa gaa acc ttc ttg aag aAG TAC Ttg tac gaa att
!      ScaI....
!
!      181 182 183 184 185 186 187 188 189 190 191 192 193 194 195
!      A   R   R   H   P   Y   F   Y   A   P   E   L   L   F   F
! 1369  gct aga aga cac cca tac ttc tac gct cca gaa ttg ttg ttc ttc
!
!      196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
!      A   K   R   Y   K   A   A   F   T   E   C   C   Q   A   A
! 1414  gct aag aga tac aag gct gct ttc acc gaa tgt tgt caa gct gct
!
!      211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
!      D   K   A   A   C   L   L   P   K   L   D   E   L   R   D
! 1459  gat aag gct gct tgt ttg ttg cca aag ttg gat gaa ttg aga gac
!
!      226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
!      E   G   K   A   S   S   A   K   Q   R   L   K   C   A   S
! 1504  gaa ggt aag gct tct tcc gct aag caa aga ttg aag tgt gct tcc
!
!      241 242 243 244 245 246 247 248 249 250 251 252 253 254 255
!      L   Q   K   F   G   E   R   A   F   K   A   W   A   V   A
! 1549  ttg caa aag ttc ggt gaa aga gct ttc aag gct tgg gct gtc gct
!
!      256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
!      R   L   S   Q   R   F   P   K   A   E   F   A   E   V   S
! 1594  aga ttg tct caa aga ttc cca aag gct gaa ttc gct gaa gtt tct
!
!      271 272 273 274 275 276 277 278 279 280 281 282 283 284 285
!      K   L   V   T   D   L   T   K   V   H   T   E   C   C   H

```

- 83 -

```

1639   aag ttg gtt act gac ttg act aag gtt 'cac act gaa tgt tgt cac
!
!       286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
!       G   D   L   L   E   C   A   D   D   R   A   D   L   A   K
1684   ggt gac ttg ttg gaa tgt gct gat gac aga gct gac ttg gct aag
!
!       301 302 303 304 305 306 307 308 309 310 311 312 313 314 315
!       Y   I   C   E   N   Q   D   S   I   S   S   K   L   K   E
1729   tac atc tgt gaa aac caa gac tct atC TCT TCc aag ttg aag gaa
!
!                               EarI....
!
!       316 317 318 319 320 321 322 323 324 325 326 327 328 329 330
!       C   C   E   K   P   L   L   E   K   S   H   C   I   A   E
1774   tgt tgt gaa aag cca ttg ttg gaa aag tct cac tgt att gct gaa
!
!       331 332 333 334 335 336 337 338 339 340 341 342 343 344 345
!       V   E   N   D   E   M   P   A   D   L   P   S   L   A   A
1819   gtt gaa aac gat gaa atg cCA GCT Gac ttg cca tct ttg gct gct
!
!                               PvuII...
!
!       346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
!       D   F   V   E   S   K   D   V   C   K   N   Y   A   E   A
1864   gac ttc gtt gaa tct aag gac gtt tgt aag aac tac gct gaa gct
!
!       361 362 363 364 365 366 367 368 369 370 371 372 373 374 375
!       K   D   V   F   L   G   M   F   L   Y   E   Y   A   R   R
1909   aag gac gtc ttc ttg ggt atg ttc ttg tac gaa tac gct aga aga
!
!       376 377 378 379 380 381 382 383 384 385 386 387 388 389 390
!       H   P   D   Y   S   V   V   L   L   L   R   L   A   K   T
1954   cac cca gac tac tcc gtt gtc ttg ttg ttg aga ttg gct aag acc
!
!       391 392 393 394 395 396 397 398 399 400 401 402 403 404 405
!       Y   E   T   T   L   E   K   C   C   A   A   A   D   P   H
1999   tac gaa act acc ttg gaa aag tgt tgt gct gct gct gac cca cac
!
!       406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
!       E   C   Y   A   K   V   F   D   E   F   K   P   L   V   E
2044   gaa tgt tac gct aag gtt ttc gat gaa ttc aag cca ttg gtc gaa
!
!       421 422 423 424 425 426 427 428 429 430 431 432 433 434 435
!       E   P   Q   N   L   I   K   Q   N   C   E   L   F   E   Q
2089   gaa cca caa aac tTG ATC Aag caa aac tgt gaa ttg ttc gaa caa
!
!                               BclI....
!
!       436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
!       L   G   E   Y   K   F   Q   N   A   L   L   V   R   Y   T
2134   ttg ggt gaa tac aag ttc caa aac gct ttg ttg gtt aga tac act
!
!       451 452 453 454 455 456 457 458 459 460 461 462 463 464 465
!       K   K   V   P   Q   V   S   T   P   T   L   V   E   V   S
2179   aag aag gtc cca caa gtc tCC Acc cca act tTG Gtt gaa gtc TCT
!
!                               XcmI.....
!
!       466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
!       R   N   L   G   K   V   G   S   K   C   C   K   H   P   E
2224   AGA aac ttg ggt aag gtc ggt tct aag tgt tgt aag cac cca gaa
!
!       481 482 483 484 485 486 487 488 489 490 491 492 493 494 495
!       A   K   R   M   P   C   A   E   D   Y   L   S   V   V   L
2269   gct aag aGA ATG Cca tgt gct gaa gat tac ttg tcc gtc gtt ttg

```

- 84 -

```

!
!           BsmI....
!
!   496 497 498 499 500 501 502 503 504 505 506 507 508 509 510
!   N   Q   L   C   V   L   H   E   K   T   P   V   S   D   R
! 2314 aac caa ttg tgt gtt ttg cac gaa aaG ACc cca GTC tct gat aga
!                                     PshAI.....
!                                     AlwNI.....
!
!   511 512 513 514 515 516 517 518 519 520 521 522 523 524 525
!   V   T   K   C   C   T   E   S   L   V   N   R   R   P   C
! 2359 gtC ACc aaG TGt tgt act gaa tct ttg GTT AAC aga aga cca tgt
!       DraIII.....
!       HpaI...
!
!   526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
!   F   S   A   L   E   V   D   E   T   Y   V   P   K   E   F
! 2404 ttc tct gct ttg gaa GTC GAC gaa act tac gtt cca aag GAA TTC
!                                     SallI...
!
!   541 542 543 544 545 546 547 548 549 550 551 552 553 554 555
!   N   A   E   T   F   T   F   H   A   D   I   C   T   L   S
! 2449 aac gct gaa act ttc acc ttc cac gct GAT ATC tgt acc ttg tcc
!                                     EcoRV..
!
!   556 557 558 559 560 561 562 563 564 565 566 567 568 569 570
!   E   K   E   R   Q   I   K   K   Q   T   A   L   V   E   L
! 2494 gaa aag gaa aga caa att aag aag caa act gct ttg gtt gaa ttg
!
!   571 572 573 574 575 576 577 578 579 580 581 582 583 584 585
!   V   K   H   K   P   K   A   T   K   E   Q   L   K   A   V
! 2539 gtc aag cac aag cca aag gct act aag gaa caa ttg aag gct gtc
!
!   586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
!   M   D   D   F   A   A   F   V   E   K   C   C   K   A   D
! 2584 atg gat gat ttc gct gct ttc gtt gaa aag tgt tgt aag gct gat
!
!   601 602 603 604 605 606 607 608 609 610 611 612 613 614 615
!   D   K   E   T   C   F   A   E   E   G   K   K   L   V   A
! 2629 gat aag gaa act tgt ttc gct gaa gaa ggt aag aag ttg gtc gct
!
!   616 617 618 619 620 621 622 623 624 625 626 627 628 629 630
!   A   S   Q   A   A   L   G   L   G   G   S   G   G   S   G
! 2674 gct tcc caa gct gCC TTA GGc tta ggt ggt tct ggt ggt tcc ggt
!                                     Bsu36I...
!
!   631 632 633 634 635 636 637 638
!   G   S   G   G   S   G   G   T
! 2719 ggt TCC GGA ggt tcc ggt GGT ACC          taa tAA GCTTa attcctatga
!       BspEI..          KpnI...          Stop Stop
!                                     HindIII (2/2)
!
! 2764 tttatgattt ttattattaa ataagTTATA Aaaaaaataa gtGTATACaa attttaaagt
!                                     PsiI...          BstZ17I
!
! 2824 gactccttagg ttttaaaacg aaaattctta ttcttgagta actctttcct gtaggtcagg
! 2884 ttgctttctc aggtatagca tgaggtcgtc cttattgacc acacctctac cgGCATGCCg
!                                     SphI..
!
! 2944 agcaaatgcc tgcaaatcgc tccccatttc acccaattgt agatatgcta actccagcaa
! 3004 tgagttgatg aatctcggtg tgtattttat gtcctcagag gacaacacct gttgtaatcg
! 3064 ttcttccaca cggatCGCGG CCGC
!                                     NotI.....

```

- 85 -

(SEQ. ID NO.: \_\_\_\_\_)

**Table 26: NotI cassette of pDB2300X2 with  
DX890(Nterm) and Cterm linker ready for second DX890**

```

!
!   1  GCGGCCGCcc gtaatgcggt atcgtgaaag cgaaaaaaaa actaacagta gataagacag
!   NotI....
!
!   61  atagacagat agagatggac gagaaacagg gggggagaaa aggggaaaag agaaggaaaag
121  aaagactcat ctatcgcaga taagacaatc aaccctcatG GCGCCTccaa ccaccatccg
!                                     NarI...
!
!   181 cactagggac caAGCGCTcg caccgttagc aacgcttgac tcacaaacca actGCCGGCt
!                                     AfeI..                               NgoMIV
!
!   241 gaaagagctt gtgcaatggg agtgccaatt caaaggagcc gaatacgtct gctcgccttt
301  taagaggctt tttgaacact gcattgcacc cgacaaatca gccactaact acgaggtcac
361  ggacacatat accaatagtt aaaaattaca tatactctat atagcacagt agtgtgataa
421  ataaaaaatt ttgccaagac ttttttaaaC TGCACccgac agatcaggtc tgtgcctact
!                                     BsgI...
!
!   481 atgcacttat gcccggggtc ccgggaggag aaaaaacgag ggctgggaaa tgtccgtgga
541  ctttaaacgc tccgggttag cagagtaGCA gggcttTCGg ctttggaat ttaggtgact
!                                     BcgI.....
!
!   601 tgttgaaaaa gcaaaatttg ggctcagtaa tgCCActgca gTGGcttatc acgccaggac
!                                     BstXI.....
!                                     PStI...
!
!   661 tgcggggagtg gcggggggcaa acacacccgc gataaagagc gcgatgaata taaaaggggg
721  ccaatgttac gtcccgttat attggagttc ttcccataca aaCTTAAGag tccaattagc
!                                     AflII.
!
!   781 ttcatcgcca ataaaaaaaaac AAGCTTaacc taattctaac aagcaaag
!                                     HindIII (1/2)
!
!   Signal sequence-----
!   1  2  3  4  5  6  7  8  9  10 11 12 13 14 15
!   M  K  W  V  F  I  V  S  I  L  F  L  F  S  S
!   829 atg aag tgg gtt ttc atc gtc tcc att ttg ttc ttg ttc tcc tct
!
!   Signal sequence-----> DX-890-----
!   16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
!   A  Y  S  R  S  L  D  K  R  E  A  C  N  L  P
!   874 gct tac tct AGA TCT ttg gat aag aga gaa gcc tgt aac ttg cca
!                                     BglII..
!                                     XbaI...(1/2)
!
!   DX890 continued-----
!   31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
!   I  V  R  G  P  C  I  A  F  F  P  R  W  A  F
!   919 att gtt aga ggt cca tgt att gct ttc ttc cca aga tgg gct ttc
!

```

- 86 -

```

!      DX890 continued-----
!      46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
!      D  A  V  K  G  K  C  V  L  F  P  Y  G  G  C
! 964   gat gct gtt aag ggt aag tgt gtt ttg ttc CCA tat ggT GGt tgt
!                                     PflMI.....
!                                     NdeI....
!
!      DX890 continued-----
!      61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
!      Q  G  N  G  N  K  F  Y  S  E  K  E  C  R  E
! 1009  caa ggt aac ggt aac aag ttc tac tct gaa aag gaa tgt aga gaa
!
!      DX890 continued---> Linker-----
!      76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
!      Y  C  G  V  P  G  G  S  G  G  S  G  G  S  G
! 1054  tac tgt ggt gtt cca ggt GGA TCC ggt ggt tcc ggt ggt tct ggt
!                                     BamHI..
!
!      Linker-----> rHA-----> to residue 679
!      91 92 93 94 95 96 97 98 99 100 101 102 103 104 105
!      G  S  G  G  D  A  H  K  S  E  V  A  H  R  F
! 1099  ggt tcc ggt ggt gac gct cac aag tcc gaa gtc gct cAC CGG Ttc
!                                     AgeI....
!
!      106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
!      K  D  L  G  E  E  N  F  K  A  L  V  L  I  A
! 1144  aag gaC CTA GGt gag gaa aac ttc aag gct ttg gtc ttg atc gct
!                                     AvrII...
!
!      121 122 123 124 125 126 127 128 129 130 131 132 133 134 135
!      F  A  Q  Y  L  Q  Q  C  P  F  E  D  H  V  K
! 1189  ttc gct caa tac ttg caa caa tgt cca ttc gaa gat CAC GTC aag
!                                     BmgBI..
!
!      136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
!      L  V  N  E  V  T  E  F  A  K  T  C  V  A  D
! 1234  ttg gtc aac gaa gtt acc gaa ttc gct aag act tgt gtt gct gac
!
!      151 152 153 154 155 156 157 158 159 160 161 162 163 164 165
!      E  S  A  E  N  C  D  K  S  L  H  T  L  F  G
! 1279  gaa tct gct gaa aac tgt gac aag tcc ttg cac acc ttg ttc ggt
!
!      166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
!      D  K  L  C  T  V  A  T  L  R  E  T  Y  G  E
! 1324  gat aag ttg tgt act gtt gct acc ttg aga gaa acc tac ggt gaa
!
!      181 182 183 184 185 186 187 188 189 190 191 192 193 194 195
!      M  A  D  C  C  A  K  Q  E  P  E  R  N  E  C
! 1369  atg gct gac tgt tgt gct aag caa gaa cca gaa aga aac gaa tgt
!
!      196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
!      F  L  Q  H  K  D  D  N  P  N  L  P  R  L  V
! 1414  ttc ttg caa cac aag gac gac aac cca aac ttg cca aga ttg gtt
!
!      211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
!      R  P  E  V  D  V  M  C  T  A  F  H  D  N  E
! 1459  aga cca gaa gtt gac gtc atg tgt act gct ttc cac gac aac gaa
!
!      226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
!      E  T  F  L  K  K  Y  L  Y  E  I  A  R  R  H
! 1504  gaa acc ttc ttg aag aAG TAC Ttg tac gaa att gct aga aga cac

```



ScaI...

		ScaI....													
1549	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
	P	Y	F	Y	A	P	E	L	L	F	F	A	K	R	Y
	cca	tac	ttc	tac	gct	cca	gaa	ttg	ttg	ttc	ttc	gct	aag	aga	tac
1594	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270
	K	A	A	F	T	E	C	C	Q	A	A	D	K	A	A
	aag	gct	gct	ttc	acc	gaa	tgt	tgt	caa	gct	gct	gat	aag	gct	gct
1639	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285
	C	L	L	P	K	L	D	E	L	R	D	E	G	K	A
	tgt	ttg	ttg	cca	aag	ttg	gat	gaa	ttg	aga	gac	gaa	ggg	aag	gct
1684	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
	S	S	A	K	Q	R	L	K	C	A	S	L	Q	K	F
	tct	tcc	gct	aag	caa	aga	ttg	aag	tgt	gct	tcc	ttg	caa	aag	ttc
1729	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315
	G	E	R	A	F	K	A	W	A	V	A	R	L	S	Q
	ggg	gaa	aga	gct	ttc	aag	gct	tgg	gct	gtc	gct	aga	ttg	tct	caa
1774	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330
	R	F	P	K	A	E	F	A	E	V	S	K	L	V	T
	aga	ttc	cca	aag	gct	gaa	ttc	gct	gaa	gtt	tct	aag	ttg	gtt	act
1819	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345
	D	L	T	K	V	H	T	E	C	C	H	G	D	L	L
	gac	ttg	act	aag	gtt	cac	act	gaa	tgt	tgt	cac	ggg	gac	ttg	ttg
1864	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
	E	C	A	D	D	R	A	D	L	A	K	Y	I	C	E
	gaa	tgt	gct	gat	gac	aga	gct	gac	ttg	gct	aag	tac	atc	tgt	gaa
1909	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
	N	Q	D	S	I	S	S	K	L	K	E	C	C	E	K
	aac	caa	gac	tct	atC	TCT	TCc	aag	ttg	aag	gaa	tgt	tgt	gaa	aag
EarI....															
1954	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390
	P	L	L	E	K	S	H	C	I	A	E	V	E	N	D
	cca	ttg	ttg	gaa	aag	tct	cac	tgt	att	gct	gaa	gtt	gaa	aac	gat
1999	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405
	E	M	P	A	D	L	P	S	L	A	A	D			

- 88 -

```

!      451 452 453 454 455 456 457 458 459 460 461 462 463 464 465
!      L   E   K   C   C   A   A   A   D   P   H   E   C   Y   A
! 2179  ttg gaa aag tgt tgt gct gct gct gac cca cac gaa tgt tac gct
!
!      466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
!      K   V   F   D   E   F   K   P   L   V   E   E   P   Q   N
! 2224  aag gtt ttc gat gaa ttc aag cca ttg gtc gaa gaa cca caa aac
!
!      481 482 483 484 485 486 487 488 489 490 491 492 493 494 495
!      L   I   K   Q   N   C   E   L   F   E   Q   L   G   E   Y
! 2269  tTG ATC Aag caa aac tgt gaa ttg ttc gaa caa ttg ggt gaa tac
!      BclI....
!
!      496 497 498 499 500 501 502 503 504 505 506 507 508 509 510
!      K   F   Q   N   A   L   L   V   R   Y   T   K   K   V   P
! 2314  aag ttc caa aac gct ttg ttg gtt aga tac act aag aag gtc cca
!
!      511 512 513 514 515 516 517 518 519 520 521 522 523 524 525
!      Q   V   S   T   P   T   L   V   E   V   S   R   N   L   G
! 2359  caa gtc tCC Acc cca act tTG Gtt gaa gtc TCT AGA aac ttg ggt
!      XcmI..... XbaI... (2/2)
!
!      526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
!      K   V   G   S   K   C   C   K   H   P   E   A   K   R   M
! 2404  aag gtc ggt tct aag tgt tgt aag cac cca gaa gct aag aGA ATG
!      BsmI....
!
!      541 542 543 544 545 546 547 548 549 550 551 552 553 554 555
!      P   C   A   E   D   Y   L   S   V   V   L   N   Q   L   C
! 2449  Cca tgt gct gaa gat tac ttg tcc gtc gtt ttg aac caa ttg tgt
!      BsmI..
!
!      556 557 558 559 560 561 562 563 564 565 566 567 568 569 570
!      V   L   H   E   K   T   P   V   S   D   R   V   T   K   C
! 2494  gtt ttg cac gaa aaG ACc cca GTC tct gat aga gtC ACc aaG TGT
!      PshAI..... DraIII.....
!      AlwNI.....
!
!      571 572 573 574 575 576 577 578 579 580 581 582 583 584 585
!      C   T   E   S   L   V   N   R   R   P   C   F   S   A   L
! 2539  tgt act gaa tct ttg GTT AAC aga aga cca tgt ttc tct gct ttg
!      HpaI...
!
!      586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
!      E   V   D   E   T   Y   V   P   K   E   F   N   A   E   T
! 2584  gaa GTC GAC gaa act tac gtt cca aag gaa ttc aac gct gaa act
!      SalI...
!
!      601 602 603 604 605 606 607 608 609 610 611 612 613 614 615
!      F   T   F   H   A   D   I   C   T   L   S   E   K   E   R
! 2629  ttc acc ttc cac gct GAT ATC tgt acc ttg tcc gaa aag gaa aga
!      EcoRV..
!
!      616 617 618 619 620 621 622 623 624 625 626 627 628 629 630
!      Q   I   K   K   Q   T   A   L   V   E   L   V   K   H   K
! 2674  caa att aag aag caa act gct ttg gtt gaa ttg gtc aag cac aag
!
!      631 632 633 634 635 636 637 638 639 640 641 642 643 644 645
!      P   K   A   T   K   E   Q   L   K   A   V   M   D   D   F
! 2719  cca aag gct act aag gaa caa ttg aag gct gtc atg gat gat ttc

```

- 89 -

```

!      646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
!      A  A  F  V  E  K  C  C  K  A  D  D  K  E  T
! 2764 gct gct ttc gtt gaa aag tgt tgt aag gct gat gat aag gaa act
!
!      661 662 663 664 665 666 667 668 669 670 671 672 673 674 675
!      C  F  A  E  E  G  K  K  L  V  A  A  S  Q  A
! 2809 tgt ttc gct gaa gaa ggt aag aag ttg gtc gct gct tcc caa gct
!
!      676 677 678 679 680 681 682 683 684 685 686 687 688 689 690
!      A  L  G  L  G  G  S  G  G  S  G  G  S  G  G
! 2854 gCC TTA GGC tta ggt ggt tct ggt ggt tcc ggt ggt TCC GGA ggt
!      Bsu36I...                               BspEI...
!
!      691 692 693 694
!      S  G  G  T
! 2899 tcc ggt GGT ACC          taa tAA GCTTa attccttatga
!      KpnI...          Stop Stop
!                               HindIII (2/2)
!
! 2932 tttatgattt ttattattaa ataagTTATA Aaaaaaataa gtGTATACaa attttaaagt
!                               PsiI...          BstZ17I
!
! 2992 gactcttagg ttttaaaacg aaaattccta ttcttgagta actctttcct gtaggtcagg
! 3052 ttgctttctc aggtatagca tgaggtcgct cttattgacc acacctctac cgGCATGCcg
!                               SphI...
!
! 3112 agcaaattgcc tgcaaattgcc tccccatttc acccaattgt agatatgcta actccagcaa
! 3172 tgagttgatg aatctcggtg tgtattttat gtcctcagag gacaacacct gttgtaatcg
! 3232 ttcttccaca cggatCGCGG CCGC
!                               NotI.....

```

(SEQ. ID NO: \_\_\_\_\_)

**Table 27: DNA to insert at BspEI/KpnI site for 2<sup>nd</sup> encoding of DX-890**

TCCGGAaggta gtgggtggctc cgggtgggtgag gcttgcaatc ttcctatcgt  
 Ccgtggccct tgcatacgct tttttcctcg ttgggccttt gacgccgtca  
 Aaggcaaattg cgtccttttt ccttacggcg gttgccaggg caatggcaat  
 AaatTTTTata gcgagaaaga gtgccgtgag tattgcggcg tcccttaata  
 aGGTACC (SEQ. ID NO: \_\_\_\_\_)

**Table 28: NotI cassette of pDB2300X3 with 2 x DX890**

```

!
! DNA sequence has SEQ ID NO: _____
! AA Sequence has SEQ ID NO:
! Enzymes that cut from 1 to 3 times.
!
! $ = DAM site, * = DCM site, & = both
!
!NotI GCggccgc          2          1    3434
!EagI Cggccg           2          2    3435
!KasI Ggcgcc           1          160
!AfeI AGCgct           1          193
!NaeI GCCggc           1          234
!NcoMIV Gccggc         1          234

```

- 90 -

!BsgI ctgcac	1	450	
!BcgI gcannnnntcg	1	568	
!BanII GRGcYc	1	620	
!PstI CTGCAG	1	636	
!AflIII Cttaag	1	763	
!HindIII Aagctt	2	801	3101
!BglII Agatct	1	883\$	
!PflMI CCANNNNNntgg	1	994	
!NdeI CATatg	1	995	
!BamHI Ggatcc	1	1072\$	
!AgeI Accggt	1	1136	
!AvrII Cctagg	1	1149	
!BmgBI CACgtc	1	1225\$	
!ScaI AGTact	1	1520	
!EarI CTCTTCNnnn	1	1923	
!PvuII CAGctg	1	2006	
!BclI Tgatca	1	2270\$	
!XcmI CCANNNNNnnnnntgg	1	2366	
!BsmI GAATGCN	1	2444	
!PshAI GACNNnngtc	1	2508	
!AlwNI CAGNNNctg	1	2513	
!DraIII CACNNNgtg	1	2529	
!HpaI GTTaaC	1	2554	
!SalI Gtcgac	1	2587	
!EcoRV GATatc	1	2644	
!Bsu36I CCtnagg	1	2855	
!BspEI Tccgga	1	2890	
!PflFI GACNnngtc	1	2980	
!Tth111I GACNnngtc	1	2980	
!Acc65I Ggtacc	1	3091	
!KpnI GGTACc	1	3091	
!PsiI TTAtaa	1	3143	
!BstZ17I GTAtac	1	3160	
!SphI GCATGc	1	3290	

!-----

1 GCGGCCCGCcc gtaatgcggt atcgtgaaag cgaaaaaaaa actaacagta gataagacag  
! NotI....

61 atagacagat agagatggac gagaaacagg ggggggagaaa aggggaaaaag agaaggaaaag  
121 aaagactcat ctatcgcaga taagacaatc aaccctcatG GCGCCTccaa ccaccatccg  
! NarI...

181 cactagggac caAGCGCTcg caccgttagc aacgcttgac tcacaaacca actGCCGGCt  
! AfeI.. NgoMIV

241 gaaagagctt gtgcaatggg agtgccaatt caaaggagcc gaatacgtct gctcgccctt  
301 taagaggctt tttgaacact gcattgcacc cgacaaatca gccactaact acgaggtcac  
361 ggacacatat accaatagtt aaaaattaca tatactctat atagcacagt agtgtgataa  
421 ataaaaaatt ttgccaagac tttttttaaC TGCACccgac agatcagggtc tgtgcctact  
! BsgI...

481 atgcacttat gcccgggggtc ccggggaggag aaaaaacgag ggctgggaaa tgtccgtgga  
541 ctttaaacgc tccgggttag cagagtaGCA gggcttTCGg ctttggaat ttaggtgact  
! BcgI.....

601 tgttgaaaaa gcaaaatttg ggctcagtaa tgCCActgca gTGGcttatc acgccaggac  
! BstXI.....  
! PstI...

661 tgcggggagt gcgggggcaa acacacccgc gataaagagc gcgatgaata taaaaggggg  
721 ccaatgttac gtcccgttat attggagttc ttcccatata aaCTTAAGag tccaattagc

- 91 -

AflIII.

781 ttcacgcgcca ataaaaaaac AAGCTTaacc taattctaac aagcaaag  
HindIII (1/2)

Signal sequence ----->

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15  
M K W V F I V S I L F L F S S  
829 atg aag tgg gtt ttc atc gtc tcc att ttg ttc ttg ttc tcc tct

Signal sequence -----> DX890, first instance -->

16 17 18 19 20 21 22 23 24 25 26 27 28 29 30  
A Y S R S L D K R E A C N L P  
874 gct tac tct AGA TCT ttg gat aag aga gaa gcc tgt aac ttg cca  
BglIII..

31 32 33 34 35 36 37 38 39 40 41 42 43 44 45  
I V R G P C I A F F P R W A F  
919 att gtt aga ggt cca tgt att gct ttc ttc cca aga tgg gct ttc

46 47 48 49 50 51 52 53 54 55 56 57 58 59 60  
D A V K G K C V L F P Y G G C  
964 gat gct gtt aag ggt aag tgt gtt ttg ttc CCA tat ggT GGt tgt  
PflMI.....  
NdeI....

61 62 63 64 65 66 67 68 69 70 71 72 73 74 75  
Q G N G N K F Y S E K E C R E  
1009 caa ggt aac ggt aac aag ttc tac tct gaa aag gaa tgt aga gaa

----DX890#1-----> ----- Linker -----

76 77 78 79 80 81 82 83 84 85 86 87 88 89 90  
Y C G V P G G S G G S G G S G  
1054 tac tgt ggt gtt cca ggt GGA TCC ggt ggt tcc ggt ggt tct ggt  
BamHI..

--- Linker ---> ----- rHA gene -----until codon 679 --->

91 92 93 94 95 96 97 98 99 100 101 102 103 104 105  
G S G G D A H K S E V A H R F  
1099 ggt tcc ggt ggt gac gct cac aag tcc gaa gtc gct cAC CGG Ttc  
AgeI....

106 107 108 109 110 111 112 113 114 115 116 117 118 119 120  
K D L G E E N F K A L V L I A  
1144 aag gaC CTA GGt gag gaa aac ttc aag gct ttg gtc ttg atc gct  
AvrII...

121 122 123 124 125 126 127 128 129 130 131 132 133 134 135  
F A Q Y L Q Q C P F E D H V K  
1189 ttc gct caa tac ttg caa caa tgt cca ttc gaa gat cac gtc aag

136 137 138 139 140 141 142 143 144 145 146 147 148 149 150  
L V N E V T E F A K T C V A D  
1234 ttg gtc aac gaa gtt acc gaa ttc gct aag act tgt gtt gct gac

151 152 153 154 155 156 157 158 159 160 161 162 163 164 165  
E S A E N C D K S L H T L F G  
1279 gaa tct gct gaa aac tgt gac aag tcc ttg cac acc ttg ttc ggt

166 167 168 169 170 171 172 173 174 175 176 177 178 179 180  
D K L C T V A T L R E T Y G E

- 92 -

1324 gat aag ttg tgt act gtt gct acc ttg aga gaa acc tac ggt gaa  
 !  
 ! 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195  
 ! M A D C C A K Q E P E R N E C  
 1369 atg gct gac tgt tgt gct aag caa gaa cca gaa aga aac gaa tgt  
 !  
 ! 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210  
 ! F L Q H K D D N P N L P R L V  
 1414 ttc ttg caa cac aag gac gac aac cca aac ttg cca aga ttg gtt  
 !  
 ! 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225  
 ! R P E V D V M C T A F H D N E  
 1459 aga cca gaa gtt gac gtc atg tgt act gct ttc cac gac aac gaa  
 !  
 ! 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240  
 ! E T F L K K Y L Y E I A R R H  
 1504 gaa acc ttc ttg aag aag tac ttg tac gaa att gct aga aga cac  
 !  
 ! 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255  
 ! P Y F Y A P E L L F F A K R Y  
 1549 cca tac ttc tac gct cca gaa ttg ttg ttc ttc gct aag aga tac  
 !  
 ! 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270  
 ! K A A F T E C C Q A A D K A A  
 1594 aag gct gct ttc acc gaa tgt tgt caa gct gct gat aag gct gct  
 !  
 ! 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285  
 ! C L L P K L D E L R D E G K A  
 1639 tgt ttg ttg cca aag ttg gat gaa ttg aga gac gaa ggt aag gct  
 !  
 ! 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300  
 ! S S A K Q R L K C A S L Q K F  
 1684 tct tcc gct aag caa aga ttg aag tgt gct tcc ttg caa aag ttc  
 !  
 ! 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315  
 ! G E R A F K A W A V A R L S Q  
 1729 ggt gaa aga gct ttc aag gct tgg gct gtc gct aga ttg tct caa  
 !  
 ! 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330  
 ! R F P K A E F A E V S K L V T  
 1774 aga ttc cca aag gct gaa ttc gct gaa gtt tct aag ttg gtt act  
 !  
 ! 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345  
 ! D L T K V H T E C C H G D L L  
 1819 gac ttg act aag gtt cac act gaa tgt tgt cac ggt gac ttg ttg  
 !  
 ! 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360  
 ! E C A D D R A D L A K Y I C E  
 1864 gaa tgt gct gat gac aga gct gac ttg gct aag tac atc tgt gaa  
 !  
 ! 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375  
 ! N Q D S I S S K L K E C C E K  
 1909 aac caa gac tct atc tct tcc aag ttg aag gaa tgt tgt gaa aag  
 !  
 ! 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390  
 ! P L L E K S H C I A E V E N D  
 1954 cca ttg ttg gaa aag tct cac tgt att gct gaa gtt gaa aac gat  
 !  
 ! 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405  
 ! E M P A D L P S L A A D F V E  
 1999 gaa atg cca gct gac ttg cca tct ttg gct gct gac ttc gtt gaa

- 93 -

```

!
!       406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
!       S   K   D   V   C   K   N   Y   A   E   A   K   D   V   F
! 2044   tct aag gac gtt tgt aag aac tac gct gaa gct aag gac gtc ttc
!
!       421 422 423 424 425 426 427 428 429 430 431 432 433 434 435
!       L   G   M   F   L   Y   E   Y   A   R   R   H   P   D   Y
! 2089   ttg ggt atg ttc ttg tac gaa tac gct aga aga cac cca gac tac
!
!       436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
!       S   V   V   L   L   L   R   L   A   K   T   Y   E   T   T
! 2134   tcc gtt gtc ttg ttg ttg aga ttg gct aag acc tac gaa act acc
!
!       451 452 453 454 455 456 457 458 459 460 461 462 463 464 465
!       L   E   K   C   C   A   A   A   D   P   H   E   C   Y   A
! 2179   ttg gaa aag tgt tgt gct gct gct gac cca cac gaa tgt tac gct
!
!       466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
!       K   V   F   D   E   F   K   P   L   V   E   E   P   Q   N
! 2224   aag gtt ttc gat gaa ttc aag cca ttg gtc gaa gaa cca caa aac
!
!       481 482 483 484 485 486 487 488 489 490 491 492 493 494 495
!       L   I   K   Q   N   C   E   L   F   E   Q   L   G   E   Y
! 2269   ttg atc aag caa aac tgt gaa ttg ttc gaa caa ttg ggt gaa tac
!
!       496 497 498 499 500 501 502 503 504 505 506 507 508 509 510
!       K   F   Q   N   A   L   L   V   R   Y   T   K   K   V   P
! 2314   aag ttc caa aac gct ttg ttg gtt aga tac act aag aag gtc cca
!
!       511 512 513 514 515 516 517 518 519 520 521 522 523 524 525
!       Q   V   S   T   P   T   L   V   E   V   S   R   N   L   G
! 2359   caa gtc tcc acc cca act ttg gtt gaa gtc tct aga aac ttg ggt
!
!       526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
!       K   V   G   S   K   C   C   K   H   P   E   A   K   R   M
! 2404   aag gtc ggt tct aag tgt tgt aag cac cca gaa gct aag aga atg
!
!       541 542 543 544 545 546 547 548 549 550 551 552 553 554 555
!       P   C   A   E   D   Y   L   S   V   V   L   N   Q   L   C
! 2449   cca tgt gct gaa gat tac ttg tcc gtc gtt ttg aac caa ttg tgt
!
!       556 557 558 559 560 561 562 563 564 565 566 567 568 569 570
!       V   L   H   E   K   T   P   V   S   D   R   V   T   K   C
! 2494   gtt ttg cac gaa aag acc cca gtc tct gat aga gtc acc aag tgt
!
!       571 572 573 574 575 576 577 578 579 580 581 582 583 584 585
!       C   T   E   S   L   V   N   R   R   P   C   F   S   A   L
! 2539   tgt act gaa tct ttg gtt aac aga aga cca tgt ttc tct gct ttg
!
!       586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
!       E   V   D   E   T   Y   V   P   K   E   F   N   A   E   T
! 2584   gaa gtc gac gaa act tac gtt cca aag gaa ttc aac gct gaa act
!
!       601 602 603 604 605 606 607 608 609 610 611 612 613 614 615
!       F   T   F   H   A   D   I   C   T   L   S   E   K   E   R
! 2629   ttc acc ttc cac gct gat atc tgt acc ttg tcc gaa aag gaa aga
!
!       616 617 618 619 620 621 622 623 624 625 626 627 628 629 630
!       Q   I   K   K   Q   T   A   L   V   E   L   V   K   H   K
! 2674   caa att aag aag caa act gct ttg gtt gaa ttg gtc aag cac aag
!

```

- 94 -

```

!      631 632 633 634 635 636 637 638 639 640 641 642 643 644 645
!      P   K   A   T   K   E   Q   L   K   A   V   M   D   D   F
2719   cca aag gct act aag gaa caa ttg aag gct gtc atg gat gat ttc
!
!      646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
!      A   A   F   V   E   K   C   C   K   A   D   D   K   E   T
2764   gct gct ttc gtt gaa aag tgt tgt aag gct gat gat aag gaa act
!
!      661 662 663 664 665 666 667 668 669 670 671 672 673 674 675
!      C   F   A   E   E   G   K   K   L   V   A   A   S   Q   A
2809   tgt ttc gct gaa gaa ggt aag aag ttg gtc gct gct tcc caa gct
!
!      Linker----->
!      676 677 678 679 680 681 682 683 684 685 686 687 688 689 690
!      A   L   G   L   G   G   S   G   G   S   G   G   S   G   G
2854   gCC TTA GGc tta ggt ggt tct ggt ggt tcc ggt ggt TCC GGA ggt
!      Bsu36I...                               BspEI..
!
!      DX-890(second encoding)----to end-->
!      691 692 693 694 695 696 697 698 699 700 701 702 703 704 705
!      S   G   G   S   G   G   E   A   C   N   L   P   I   V   R
2899   agt ggt ggc tcc ggt ggt gag gct tgc aat ctt cct atc gtc cgt
!
!      706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
!      G   P   C   I   A   F   F   P   R   W   A   F   D   A   V
2944   ggc cct tgc atc gcc ttt ttt cct cgt tgg gcc ttt gac gcc gtc
!
!      721 722 723 724 725 726 727 728 729 730 731 732 733 734 735
!      K   G   K   C   V   L   F   P   Y   G   G   C   Q   G   N
2989   aaa ggc aaa tgc gtc ctt ttt cct tac ggc ggt tgc cag ggc aat
!
!      736 737 738 739 740 741 742 743 744 745 746 747 748 749 750
!      G   N   K   F   Y   S   E   K   E   C   R   E   Y   C   G
3034   ggc aat aaa ttt tat agc gag aaa gag tgc cgt gag tat tgc ggc
!
!      751 752
!      V   P
3079   gtc cct taa taa          GGT ACC          taa tAA GCTTa attcctatga
!      KpnI...                      Stop Stop
!      HindIII(2/2)
!
3118   tttatgattt ttattattaa ataagTTATA Aaaaaaataa gtGTATACaa attttaagat
!      PsiI...                      BstZ17I
!
3178   gactcttagg ttttaaaacg aaaattctta ttcttgagta actctttcct gtaggtcagg
3238   ttgctttctc aggtatagca tgagggtcgt cttattgacc acacctctac cgGCATGCcg
!      SphI..
!
3298   agcaaatagcc tgcaaatacgc tccccatttc acccaattgt agatatgcta actccagcaa
3358   tgagttgatg aatctcggtg tgtattttat gtcctcagag gacaacacct gttgtaatcg
3418   ttcttccaca cggatCGCGG CCGC
!      NotI.....

```

(SEQ. ID NO: \_\_\_\_\_).

**Table 29: AA sequence of DX890::(GGS)4GG::HA::(GGS)4GG::DX890**

EACNLPIVRG PCIAFFPRWA FDAVKGKCVL FPYGGCQGNG NKFYSEKECR  
 EYCGVPGGSG GSGSGSGSGG DAHKSEVAHR FKDLGEENFK ALVLIAFAQY



- 95 -

LQQCPFEDHV KLVNEVTEFA KTCVADESAD NCDKSLHTLF GDKLCTVATL  
 RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV DVMCTAFHDN  
 EETFLKKYLY EIARRHPYFY APELLFFAKR YKAAFTECCQ AADKAACLLP  
 KLDELRLDEGK ASSAKQRLKC ASLQKFGERA FKAWAVARLS QRFPKAEFAE  
 VSKLVTDLTK VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE  
 KPLLEKSHCI AEVENDEMPA DLPSLAADFV ESKDVCKNYA EAKDVFLGMF  
 LYEYARRHPD YSVVLLRLA KTYETTLEKC CAAADPHECY AKVFDEFKPL  
 VEEPQNLIKQ NCELFEQLGE YKFQNALLRV YTKKVPQVST PTLVEVSRNL  
 GKVGSKCCKH PEAKRMPCAE DYLSVVLNQL CVLHEKTPVS DRVTKCCTES  
 LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEKE RQIKKQTALV  
 ELVKHKPKAT KEQLKAVMDD FAAFVEKCKK ADDKETCF AE EGKKLVAAASQ  
 AALGLGGSGG SGGSGGSGGS GGEACNLPIV RGPCIAFFPR WAFDAVKGKC  
 VLFPYGGCQG NGNKFYSEKE CREYCGVP (SEQ ID NO:\_\_\_\_)

**Table 30: DNA sequence of the N-terminal *Bgl*III-*Bam*HI DX-1000 cDNA**

AGA TCT TTG GAT AAG AGA  
 gag gct atg cat tcc ttc tgc gcc ttc aag  
 gct gag act ggt cct tgt aga gct agg ttc  
 gac cgt tgg ttc ttc aac atc ttc acg cgt  
 cag tgc gag gaa ttc att tac ggt ggt tgt  
 gaa ggt aac cag aac cgg ttc gaa tct cta  
 gag gaa tgt aag aag atg tgc act cgt gac  
 GGA TCC (SEQ ID NO:\_\_\_\_)

**Table 31: AA sequence of DX1000:::(GGG)4GG:::HA**

EAMHSFCAFK AETGPCRARF DRWFFNIFTR QCEEFIYGGC EGNQNRFESL  
 EECKKMCTRD GGSGGSGGSG GSGGDAHKSE VAHRFKDLGE ENFKALVLIA  
 FAQYLQQCPF EDHVKLVNEV TEFKATCVAD ESAENC DKSL HTLFGDKLCT  
 VATLRETYGE MADCCAKQEP ERNECFLQHK DDNP NLPR LV RPEVDVMCTA  
 FHDNEETFLK KYLYEIARRH PYFYAPEL LF FAKRYKAAFT ECCQAADKAA  
 CLLPKLDELRL DEGKASSAKQ RLK CASLQKF GERAFKAWAV ARLSQRFPKA  
 EFAEVSKLVT DLTKVHTECC HGDLL ECADD RADLAKYICE NQDSISSKLK  
 ECCEKPLLEK SHCIAEVEND EMPADLPSLA ADFVESKDVC KNYAEAKDVF  
 LGMFLYEYAR RHPDYSVLL LRLAKTYETT LEKCCAAADP HECYAKVFDE  
 FKPLVEEPQN LIKQNC ELF QLG EYKFQNA LLVRYTKKVP QVSTPTLVEV  
 SRNLGKVGSK CCKHPEAKRM PCAEDYLSVV LNQLCVLHEK TPVSDRVTKC  
 CTESLVNRRP CFSALVDET YVPKEFNAET FTFHADICTL SEKERQIKKQ  
 TALVELVKHK PKATKEH (SEQ ID NO:\_\_\_\_)

- 96 -

**Table 32: DNA sequence of the N-terminal *BspEI-KpnI* DX-88 cDNA-2<sup>nd</sup> encoding**

TCC GGA ggt agt ggt ggc tcc ggt ggt  
 GA<sub>g</sub> GC<sub>c</sub> ATG CAT  
 TCT TTC TGT GCT TTC AAG GCT GAC GAC GGT  
 CCG TGC AGA GCT GCT CAC CCA AGA TGG TTC  
 TTC AAC ATC TTC ACG CGA CAA TGC GAG GAG  
 TTC ATC TAC GGT GGT TGT GAG GGT AAC CAA  
 AAC AGA TTC GAG TCT CTA GAG GAG TGT AAG  
 AAG ATG TGT ACT AGA GAC GGT taa taa GGT ACC (SEQ ID NO:\_\_\_\_)

**Table 33: AA sequence of DPI14::HSA**

EAVREVCSEQ AETGPCIAFF PRWYFDVTEG KCAPFFYGGC GGNRNNFDTE  
 EYCMVAVCGSA GSGSGSGSGS GSGGDAHKSE VAHRFKDLGE ENFKALVLIA  
 FAQYLQQCPF EDHVKL VNEV TEFKATCVAD ESAENCDKSL HTLFGDKLCT  
 VATLRETYGE MADCCAKQEP ERNECF LQHK DDNP NL PRLV RPEVDVMCTA  
 FHDNEETFLK KYLYEIARRH PYFYAPELLF FAKRYKAAFT ECCQAADKAA  
 CLLPKLDEL R DEGKASSAKQ RLK CASLQKF GERA FKAWAV ARLSQRFPKA  
 EFAEVSKLVT DLTKVHTECC HGDLL ECADD RADLAKYICE NQDSISSKLK  
 ECCEKPLLEK SHCIAEVEND EMPADLPSLA ADFVESKDVC KNYAEAKDVF  
 LGMFLYFYAR RHPDYSV VLL LRLAKTYETT LEKCCAAADP HECYAKVFDE  
 FKPLVEEPQN LIKQNC ELF E QLGEYKFQNA LLVRYTKKVP QVSTPTLVEV  
 SRNLGKVGSK CCKHPEAKRM PCAEDYLSVV LNQLCVLHEK TPVSDRVTKC  
 CTESLVNRRP CFSALEVDET YVPKEFNAET FTFHADICTL SEKERQIKKQ  
 TALVELVKHK PKATKEH (SEQ ID NO:\_\_\_\_)

**CLAIMS****WE CLAIM:**

1. An albumin fusion protein comprising a Kunitz domain peptide or a fragment or variant thereof, and albumin, or a fragment or variant thereof.
2. The albumin fusion protein according to claim 1, wherein the Kunitz domain peptide or a fragment or variant thereof has a functional activity.
3. The albumin fusion protein according to claim 2, wherein the functional activity comprises inhibiting serine proteases.
4. The albumin fusion protein according to claim 2, wherein the functional activity comprises inhibiting plasmin.
5. The albumin fusion protein according to claim 2, wherein the functional activity comprises inhibiting human neutrophil elastase.
6. The albumin fusion protein according to claim 2, wherein the functional activity comprises inhibiting kallikrein.
7. The albumin fusion protein according to claim 1 comprising DX-890 or a fragment or variant thereof and albumin or a fragment or variant thereof.
8. The albumin fusion protein according to claim 1 comprising DPI-14 or a fragment or variant thereof and albumin or a fragment or variant thereof.
9. The albumin fusion protein according to claim 1 comprising DX-88 or a fragment or variant thereof and albumin or a fragment or variant thereof.
10. The albumin fusion protein according to claim 1 comprising DX-1000 or a fragment or variant thereof and albumin or a fragment or variant thereof.
11. The albumin fusion protein according to claim 1 wherein the albumin fusion protein comprises at least two Kunitz domain fusion peptides or fragments or variants thereof.

- 98 -

12. The albumin fusion protein according to claim 11, wherein each of the at least two Kunitz domain fusion peptides or fragments or variants thereof has a functional activity.

13. The albumin fusion protein according to claim 12, wherein the functional activity of one of the at least two Kunitz domain fusion peptides comprises inhibiting serine proteases.

14. The albumin fusion protein according to claim 12, wherein the functional activity of one of the at least two Kunitz domain fusion peptides comprises inhibiting plasmin.

15. The albumin fusion protein according to claim 12, wherein the functional activity of one of the at least two Kunitz domain fusion peptides comprises inhibiting human neutrophil elastase.

16. The albumin fusion protein according to claim 12, wherein the functional activity of one of the at least two Kunitz domain fusion peptides comprises inhibiting kallikrein.

17. The albumin fusion protein according to claim 11 wherein at least two of the Kunitz domain peptides or fragments or variants thereof have different amino acid sequences.

18. The albumin fusion protein of claim 1 comprising at least one fragment or variant of a peptide selected from the group consisting of DX-890, DX-88, DX-1000, and DPI-14 and albumin or a fragment or variant thereof, and wherein said albumin fragment or variant has albumin activity and said peptide fragment or variant has a functional activity.

19. The albumin fusion protein according to claim 1, wherein said albumin activity has the ability to prolong the *in vivo* half-life of a peptide selected from the group consisting of DX-890, DX-88, DX-1000, and DPI-14, or a fragment or variant thereof, compared to the *in vivo* half-life of the peptide or a fragment or variant thereof in an unfused state.

20. The albumin fusion protein according to claim 1, further comprising one or more additional albumin moieties.

- 99 -

21. The albumin fusion protein according to claim 1, wherein the albumin fusion protein comprises one or more moieties selected from the group consisting of DX-890, DX-88, DX-1000, and DPI-14, or fragments or variants thereof, or one or more additional albumin moieties.

22. The albumin fusion protein according to claim 1, wherein said fusion protein further comprises a chemical moiety.

23. The albumin fusion protein according to claim 1, wherein the Kunitz domain peptide, or fragment or variant thereof, is fused to the N-terminus of albumin or to the N-terminus of the fragment or variant of albumin.

24. The albumin fusion protein according to claim 23, wherein the Kunitz domain peptide comprises DX-890, DPI-14, DX-88, or DX-1000.

25. The albumin fusion protein of claim 1, wherein the Kunitz domain peptide or fragment of variant thereof, is fused to the C-terminus of albumin, or the C-terminus of the fragment or variant of albumin.

26. The albumin fusion protein according to claim 24, wherein the Kunitz domain peptide comprises DX-890, DPI-14, DX-88, or DX-1000.

27. The albumin fusion protein according to claim 1, wherein said Kunitz domain peptide comprises a first peptide, or fragment or variant thereof, and a second peptide, or fragment or variant thereof, and wherein said peptide, or fragment or variant thereof, is different from said second peptide, or fragment or variant thereof.

28. The albumin fusion protein according to claim 27, wherein said first peptide, or fragment or variant thereof, and said second peptide, or fragment or variant thereof is chosen from the group consisting of DX-890, DX-88, DX-1000, and DPI-14.

29. The albumin fusion protein according to claim 1, wherein the Kunitz domain peptide, or fragment or variant thereof, is separated from the albumin or the fragment or variant of albumin by a linker.

- 100 -

30. The albumin fusion protein according to claim 1, wherein the albumin fusion protein comprises the following formula:

R2-R1; R1-R2; R2-R1-R2; R2-L-R1-L-R2; R1-L-R2; R2-L-R1; or R1-L-R2-L-R1,

wherein R1 is at least one peptide selected from the group consisting of DX-890, DX-88, DX-1000, and DPI-14, or a fragment or variant thereof, L is a peptide linker, and R2 is albumin.

31. The albumin fusion protein according to claim 1, wherein the *in vitro* biological activity of the Kunitz domain peptide, or fragment or variant thereof, fused to albumin, or fragment or variant thereof, is greater than the *in vitro* biological activity of the Kunitz domain peptide, or fragment or variant thereof, in an unfused state.

32. The albumin fusion protein according to claim 1, wherein the solubility of the Kunitz domain peptide, or fragment or variant thereof, fused to albumin, or fragment or variant thereof, is greater than the solubility of the Kunitz domain peptide, or fragment or variant thereof, in an unfused state that has been subjected to the same storage, handling or physiological conditions.

33. The albumin fusion protein according to claim 30, wherein the *in vivo* biological activity of the at least one peptide, or fragment or variant thereof, fused to albumin, or fragment or variant thereof, is greater than the *in vivo* biological activity of the at least one peptide, or fragment or variant thereof, in an unfused state.

34. The albumin fusion protein according to claim 1, wherein the albumin fusion protein is non-glycosylated.

35. The albumin fusion protein according to claim 1, wherein the albumin fusion protein is expressed in yeast.

36. The albumin fusion protein according to claim 35, wherein the yeast is glycosylation deficient.

- 101 -

37. The albumin fusion protein according to claim 36 wherein the yeast is protease deficient.

38. The albumin fusion protein according to claim 1, wherein the albumin fusion protein is expressed by a mammalian cell.

39. The albumin fusion protein according to claim 38, wherein the albumin fusion protein is expressed by a mammalian cell in culture.

40. A composition comprising the albumin fusion protein of any one of claims 1-39 and a pharmaceutically acceptable carrier.

41. A method of treating a disease or disorder in a patient, comprising the step of administering the albumin fusion protein of claim 1.

42. A method of treating a patient with cystic fibrosis or a cystic fibrosis-related disease or disorder that is modulated by DX-890 and/or DPI-14, comprising the step of administering an effective amount of the albumin fusion protein of claim 1, wherein said Kunitz domain peptide is DX-890 or DPI-14, or a fragment or variant thereof.

43. A method of extending the *in vivo* half-life of DX-890 and/or DPI-14, or a fragment or variant thereof, comprising the step of fusing the DX-890 and/or DPI-14, or fragment or variant thereof, to albumin or a fragment or variant of albumin sufficient to extend the *in vivo* half-life of the DX-890 and/or DPI-14, or fragment or variant thereof, compared to the *in vivo* half-life of the DX-890 and/or DPI-14, or fragment or variant thereof, in an unfused state.

44. A method of treating a patient with hereditary angioedema or a hereditary angioedema-related disease or disorder that is modulated by DX-88, comprising the step of administering an effective amount of the albumin fusion protein of claim 1, wherein said Kunitz domain peptide is DX-88, or a fragment or variant thereof.

45. A method of extending the *in vivo* half-life of DX-88, or a fragment or variant thereof, comprising the step of fusing the DX-88, or fragment or variant thereof, to albumin or a fragment or variant of albumin sufficient to extend the *in vivo* half-life of the DX-88, or

- 102 -

fragment or variant thereof, compared to the *in vivo* half-life of the DX-88, or fragment or variant thereof, in an unfused state.

46. A method of treating a patient with cancer, a cancer-related disease, bleeding, or disorder that is modulated by DX-1000, comprising the step of administering an effective amount of the albumin fusion protein of claim 1, wherein said Kunitz domain peptide is DX-1000, or a fragment or variant thereof.

47. A method of extending the *in vivo* half-life of DX-1000, or a fragment or variant thereof, comprising the step of fusing the DX-1000, or fragment or variant thereof, to albumin or a fragment or variant of albumin sufficient to extend the *in vivo* half-life of the DX-1000, or fragment or variant thereof, compared to the *in vivo* half-life of the DX-1000, or fragment or variant thereof, in an unfused state.

48. A nucleic acid molecule comprising a polynucleotide sequence encoding the albumin fusion protein of claim 1.

49. A vector comprising the nucleic acid molecule of claim 48.

50. A host cell comprising the nucleic acid molecule of claim 48.

51. A pharmaceutical composition comprising an effective amount of the albumin fusion protein of claim 1 and a pharmaceutically acceptable carrier or excipient.

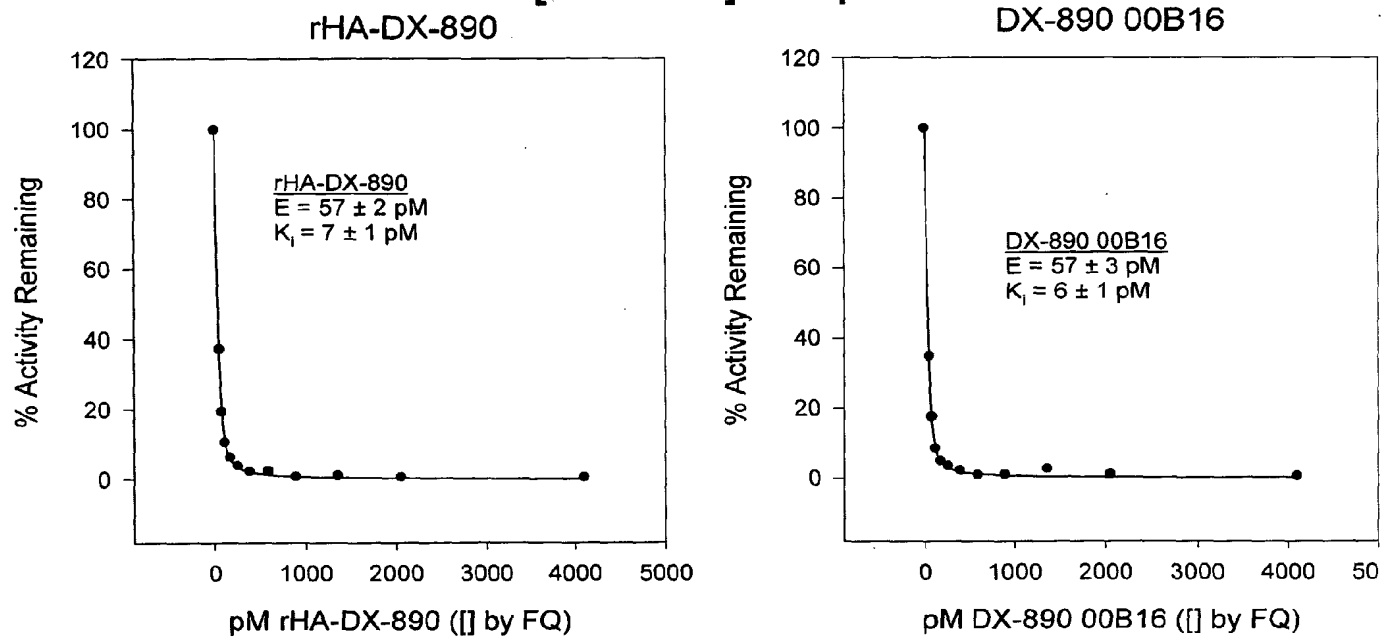
52. A method for manufacturing a albumin fusion protein of claim 1, the method comprising:

- (a) providing a nucleic acid comprising a nucleotide sequence encoding the albumin fusion protein expressible in an organism;
- (b) expressing the nucleic acid in the organism to form an albumin fusion protein; and
- (c) purifying the albumin fusion protein.

53. The method of claim 52 wherein the albumin fusion protein comprises DX-890 and/or DPI-14 albumin fusion expressed in a glycosylation deficient yeast strain.



- 1/6 -

**rHA-DX-890 Batch 1743#09  $K_i$  Determination****[HNE] = 100 pM****[substrate] = 25  $\mu$ M****FIGURE 1**

- 2/6 -

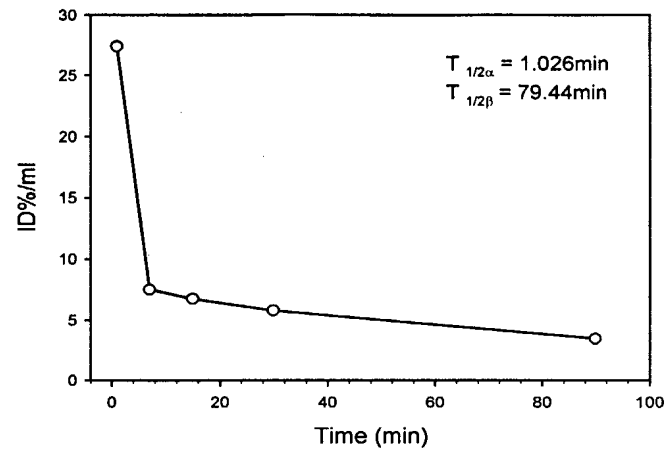
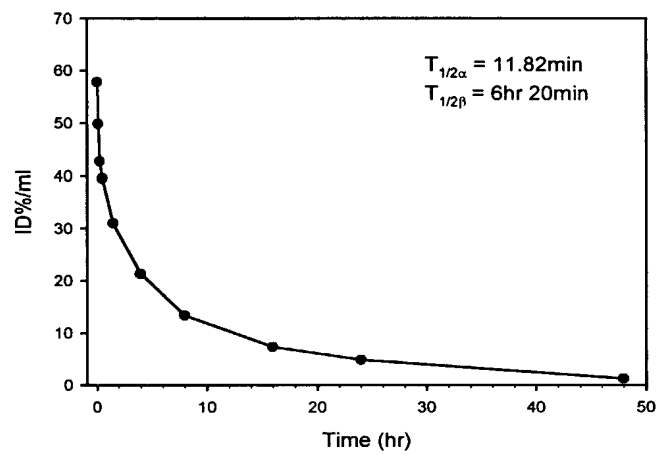
Plasma Clearance of  $^{125}\text{I}$  - DX890Plasma Clearance of  $^{125}\text{I}$ -HSA-DX890

FIGURE 2

- 3/6 -

31-2462 07/23/2002

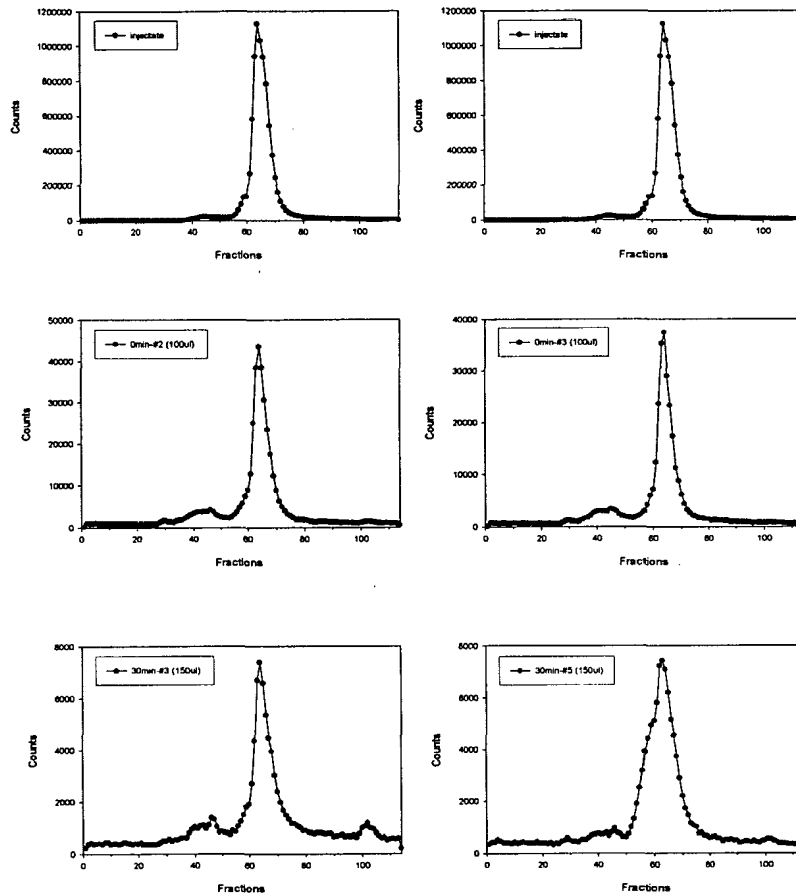
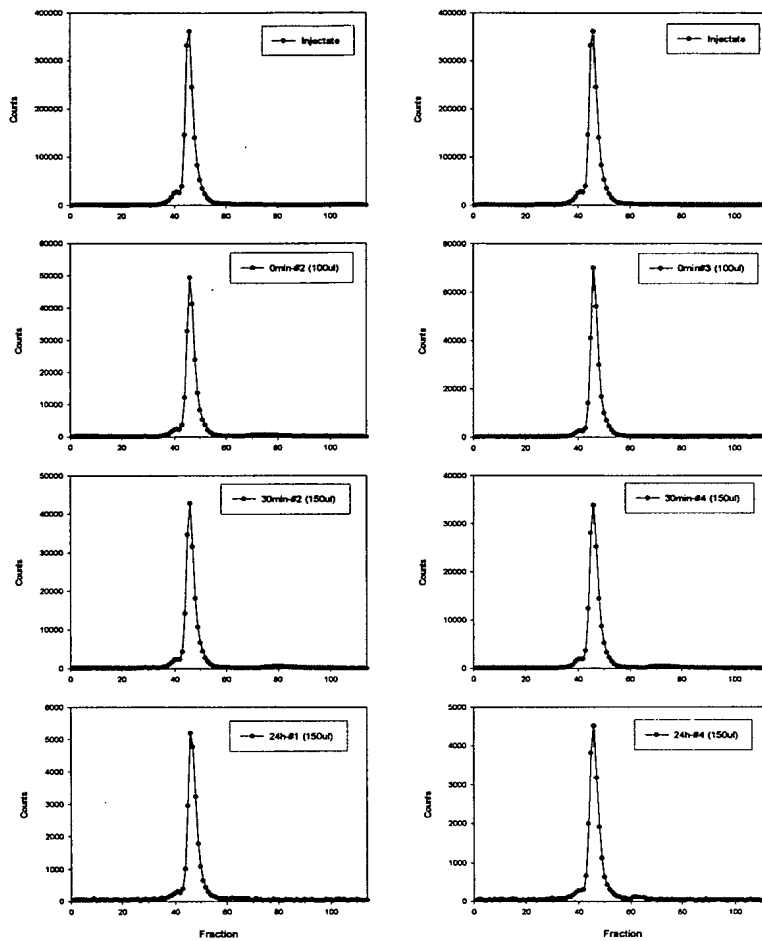
 $^{125}$ I-DX890 in Normal Mouse Plasma on SE-HPLC Superose-12

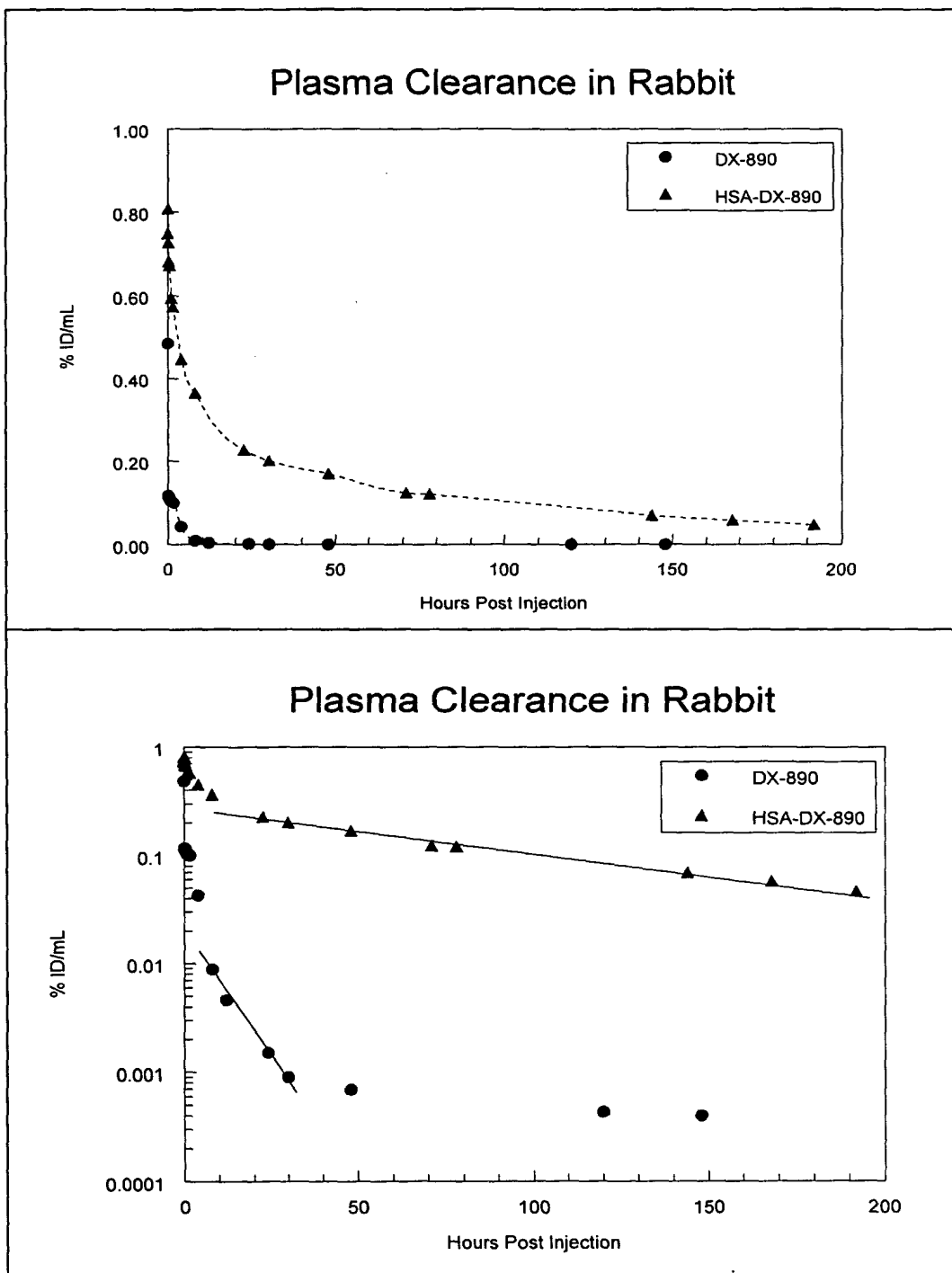
FIGURE 3

- 4/6 -

SE-HPLC(Superose-12) Profiles of  $^{125}$ I-HSA-DX890 in Normal Mouse Plasma  
31-2477  
08/14 -08/16/02

**FIGURE 4**

- 5/6 -

**Plasma Clearance of  $^{125}\text{I}$  Labeled DX-890 and HSA-DX-890 in Rabbits****FIGURE 5**

- 6/6 -

## SEC Analysis of Rabbit Plasma Samples

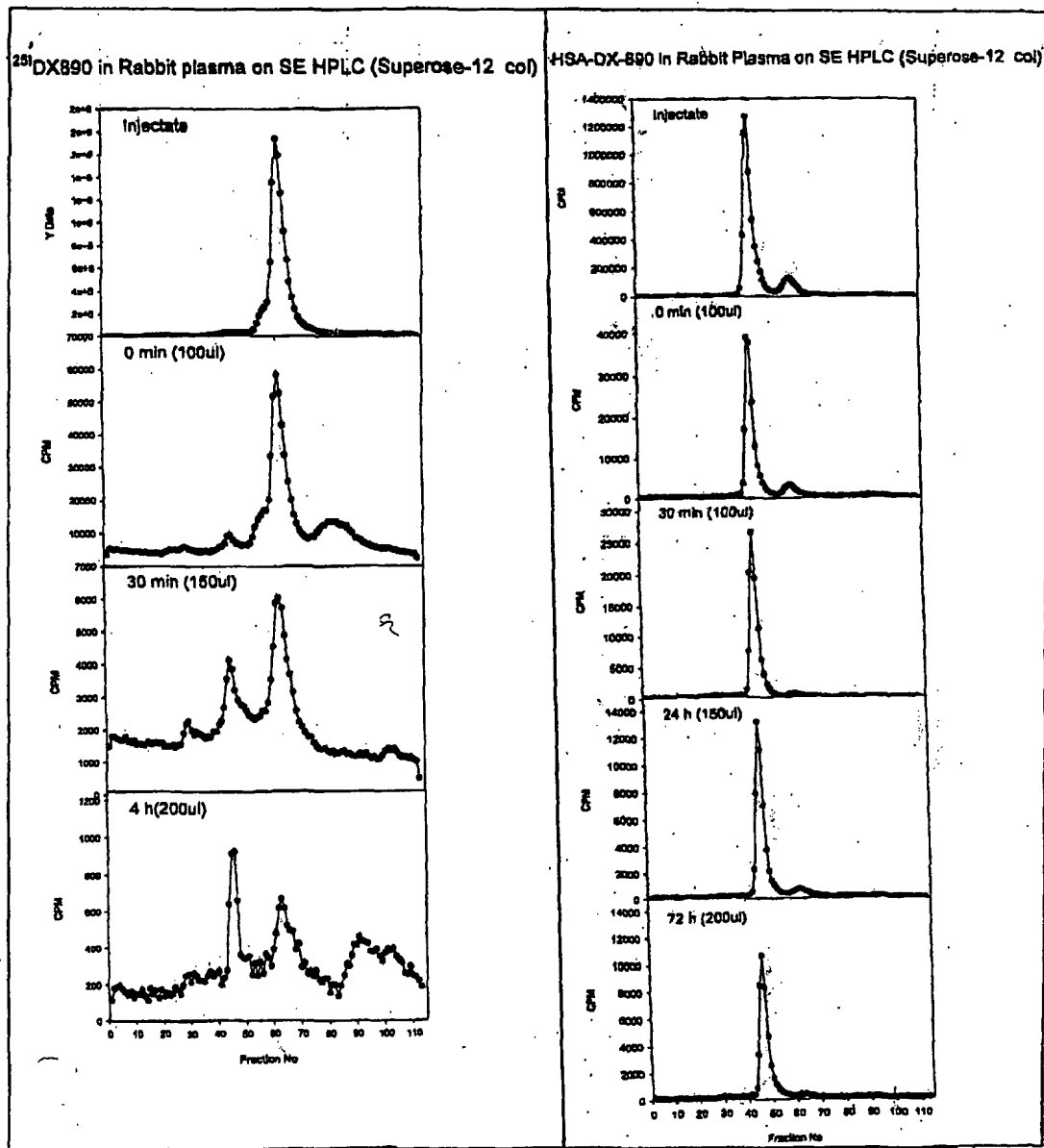


FIGURE 6